BEBR

Kalman Filter Estimates of
the MMI Cash-Futures Spread
On October 19 and 20, 1987

Gilbert W. Bassett, Jr.
Virginia G. France
Stanley R. Pliska

BEBR

Kalman Filter Estimates of the MMI Cash-Futures
Spread on October 19 and 20, 1987*

Gilbert W. Bassett, Jr.
University of Illinois at Chicago

Virginia G. France, Assistant Professor

Stanley R. Pliska
University of Illinois at Chicago

Department of Finance

ABSTRACT

Kalman Filtering is proposed as a method for estimating the value of a
stock index for periods when some of the individual stocks are not trad-
ing. A by-product of this method is a confidence interval for the esti-
mate. The method is applied to the Major Market Index during the two
principal days of the 1987 stock market crash. The resulting spread was
generally closer to zero than the widely reported spread, which is based
on the conventional, last-trade method for computing the cash index.

1. Introduction

The value of a stock index and thus a cash-futures spread can be known precisely only when the constituent stocks are continuously traded. If there are gaps in trading the valuation of the index will require estimates of the current value of nontrading securities, and the current value of the cash index will not be known with certainty. The amount of uncertainty will depend on the proportion of stocks in the index which are not trading, the length of time between trades, and the correlation between trading and nontrading stocks.

In this paper we propose Kalman filtering methods for estimating the value of a stock index when there are discontinuities in trading of the underlying stocks. We also propose an interval estimate to represent the range of possible values for both the individual nontrading stocks and the stock index during those periods when there are gaps in trading.

The method is applied to the valuation of the Major Market Index, which consists of the twenty stocks listed in Table 1. The valuation method is important for the MMI futures contract, which is traded on the Chicago Board of Trade, because its value is determined solely by the value of the underlying securities; at expiration there is cash settlement based on the prices of the stocks in the index.

When the MMI stocks are trading continuously the cash value of the futures contract can be approximated well by the most recent transaction prices of the underlying securities, and arbitrage can be expected to keep the futures price very close to the known cash value (small differences between the cash and futures prices will exist because of the costs of

carrying a portfolio of MMI stocks and because of the absence of dividends to those holding the futures contract). When there are gaps in trading, however, the cash value must be estimated without complete information on the values of the constituent stocks. There will then be uncertainty regarding the appropriate value of the futures. We propose to represent this uncertainty with an interval estimate of the current value of the cash index.

Our methods are illustrated by analyzing the MMI on Monday and Tuesday, October 19 and 20, 1987. This is a subject of independent interest since Monday was the day on which security values fell by more than 20%. Figures 1 and 2 show the futures and cash values for Monday and Tuesday.

The large spreads that are apparent in Figures 1 and 2 may be misleading since the cash values are based on the the most recently traded prices of the stocks in the index. Since prices were changing rapidly on the 19th and 20th and since there were gaps in trading, it is possible that the spread between the futures and cash values was largely due to nonsynchronous trading. In other words, better methods for estimating the cash index when component stocks were not trading may have led to spreads that were closer to zero.

In this paper transaction data and Kalman filtering are used to derive minute-by-minute point and interval estimates for the cash value of the index. This complements the previous studies which have attempted to estimate the value of stock index futures on the days around the crash; see Edwards(1989) for a review of the government reports on the crash and see Bassett, France and Pliska (1989) for our previous analysis of the cash-futures spread on Monday.

The emphasis of previous work has been on deriving point estimates of

2

the cash value of the futures contract during the times on Monday and Tuesday when stocks were not trading. The estimates for a given time period were often derived using information on both prior and subsequent prices. This differs from the Kalman filter methods considered here. The Kalman filter estimates can be implemented in real-time since they only require prior price information. The method also provides convenient interval estimates for representing the uncertainty of the cash index value when there are nontrading stocks.

The focus of our previous study was on estimating (with other, non-Kalman methods) the cash value of the MMI on Monday, the day of the crash. The estimation problem for the next day, however, is in some respects more difficult and interesting. After the opening on Monday there were (with only a few exceptions) no large gaps in trading; stocks traded almost continuously despite the steep drop in prices. With few trading gaps the estimates of the cash value using the Kalman filter are quite similar to the cash value based on the most recent trades.

The situation for Tuesday is very different. Stocks in the MMI and the market as a whole opened nearly on time and at levels well above the Monday close. At about 10:00 (Chicago time) large sell orders started to accumulate, many stocks ceased trading (for as long as two hours), and there were reports that the NYSE was going to suspend trading. Anticipation of the halt in trading of NYSE stocks led the Chicago Mercantile Exchange to suspend trading in the S&P 500 contract; the MMI futures contract however stayed open throughout Tuesday.

An indication of the general situation on Tuesday morning can be seen from the transaction data for the twenty MMI stocks. Figures 3-22 show prices when transactions occurred as well as confidence intervals for the Kalman estimated stock values; the confidence intervals will be discussed

3

in a later section. The trading gaps can be identified in the figures by noting the time intervals when only two lines are present.

The next section describes the Kalman filtering model. This model involves some statistical estimates which are discussed in Section 3. The application of the model to MMI data on Monday and Tuesday is presented in Section 4. Conclusions, other applications of the Kalman filter approach, and extensions of the analysis are discussed in the last section.

2. The Kalman Filtering Model

We use a discrete time model, where each period corresponds to one minute. Let $X_t \in \mathbb{R}^{20}$ be a column vector representing the time-t values of the 20 MMI stocks. We then assume

$$X_{t+1} = X_t + W_t$$

for all t, where $\{W_t\}$ is an iid sequence of random variables with $E[W_t] = 0$ and $E[W_t W_t'] = Q$. Here Q is a covariance matrix that needs to be specified; its estimation will be described in Section 3.

Now suppose during period t the stocks j_1, j_2, .., and j_k trade at prices $Y_t(1)$, $Y_t(2)$, .., and $Y_t(k)$, respectively, whereas the other 20 − k stocks do not trade. Let the column vector $Y_t \in \mathbb{R}^k$ denote these observations. Let B_t denote the k × 20 matrix defined by

$$B_t(i, j) = \begin{cases} 1, & j = j_1, \\ 0, & j \neq j_1. \end{cases}$$

We then assume

$$Y_t = B_t X_t + Z_t$$

for all t, where $\{Z_t\}$ is a sequence of independent random variables with

4

$E[Z_t] = 0$ and $E[Z_t Z_t'] = H_t$. The random variable $Z_t \in \mathbb{R}^k$ should be interpreted as noise; its role and the specification of its covariance matrix H_t will be discussed more fully in the next section

Now given a history Y_1, Y_2, .., Y_t of observations one seeks an optimal estimate $\hat{X}_t$ of the vector X_t of stock values. The Kalman filtering algorithm is ideally suited for this purpose, for it derives the linear, least squares estimate in a relatively simple, recursive manner. There are many references on this subject, such as Davis (1977) and Krishnan (1984), so we will not elaborate on the underlying theory.

To see how one iteration of the algorithm works for our stock index application, suppose $\hat{X}_{t-1}$ and C_{t-1} are specified, where C_{t-1} is the covariance matrix associated with the estimate $\hat{X}_{t-1}$. Since $\hat{X}_{t-1}$ is also a forecast of X_t, the "forecast error" after observing Y_t is the column vector

$$V_t \equiv Y_t - B_t \hat{X}_{t-1}.$$

Associated with V_t is another $(k \times k)$ covariance matrix

$$F_t \equiv B_t (C_{t-1} + Q) B_t' + H_t.$$

The estimate of X_t is then given by

$$\hat{X}_t = \hat{X}_{t-1} + (C_{t-1} + Q) B_t' F_t^{-1} V_t.$$

Finally, the covariance matrix associated with $\hat{X}_t$ is given by

$$C_t = C_{t-1} + Q - (C_{t-1} + Q) B_t' F_t^{-1} B_t (C_{t-1} + Q).$$

Knowing $\hat{X}_t$, the estimate $\hat{I}_t$ of the Major Market Index for period t is simply given by

$$\hat{I}_t = \hat{X}'_t 1/d,$$

where here $1 \in \mathbb{R}^{20}$ is a column vector of ones and the scalar d is the MMI divisor. Using the same notation, the variance associated with the estimate $\hat{I}_t$ is given by $1'C_t 1/d^2$.

3. The Data and Parameter Estimation

The data base for stocks consists of minute-by-minute prices for each of the twenty MMI securities. These data were extracted from transaction data, which, in turn, were obtained from the Francis Emory Fitch Company. The price at each minute was taken as the price of the last listed transaction during that minute; if there was no trade then the price was set equal to zero. Transaction data on the MMI futures contract was provided by the CBOT. The price of the futures for a minute was taken as the price of the first reported trade during that minute. The data are reported according to Chicago time from 8:30 to 3:00 for Friday, October 16, Monday, October 19, and Tuesday, October 20, 1987.

The Friday data were used for parameter estimation. To estimate the covariance matrix Q we first filled in any trading gaps with the price at the last trade, restricting attention from 8:47, when the last MMI stock opened, until 3:00. With observations then for 374 minutes, we next constructed the 373 one-minute price differences (that is, 373 20-component, first difference vectors). Finally we applied a GAUSS utility to estimate the minute-by-minute covariance matrix. The resulting 20 by 20 matrix is too large to report here, but the standard deviations of first differences for each of the 20 MMI stocks are listed in Table 1.

The covariance matrix H was arbitrarily chosen to be equal to the scalar .005 multiplied by the identity matrix; this selection is discussed

in considerable detail at the end of this section.

In order to apply the Kalman algorithm to Monday trading it remained to specify the estimate $\hat{X}_0$ of the Monday 8:30 stock values and the associated covariance matrix C_0. We took for these values the estimate of the stock values and the associated covariance matrix for the last minute on Friday after applying the Kalman filtering algorithm to all of Friday with $\hat{X}_0$ equal to the opening prices on Friday and with C_0 for Friday equal to Q. While this choice of $\hat{X}_0$ and C_0 for Friday is arbitrary, the resulting values of $\hat{X}_0$ and C_0 for Monday are not sensitive to this choice.

The values of $\hat{X}_0$ and C_0 for Tuesday were taken to be the corresponding values for the last minute on Monday, with one adjustment. A complication was presented by the fact that Eastman Kodak had a 3 for 2 stock split between Monday and Tuesday, thereby causing the MMI divisor to change from 3.18322 to 3.12165. Since Eastman Kodak did not open until 9:55 on Tuesday, the customary cash index before 9:55 on Tuesday was computed using the last Monday trade of Eastman Kodak and the old divisor (one computes the same value by taking two-thirds of the last Monday Eastman Kodak trade and using the new divisor). But for $\hat{X}_0$ on Tuesday we took the estimate for the last minute on Monday, replacing the Eastman Kodak component by two-thirds of its closing Monday value. Consequently, the Kalman estimate of the MMI cash index was computed with the new divisor from 8:30 Tuesday onward.

The covariance matrix C_0 for Tuesday was taken to be identical to the corresponding matrix for the last minute on Monday. It would have been desirable to modify this matrix in order to reflect the reduced volatility of Eastman Kodak after the split. But the impact of such an adjustment is expected to be slight, so we chose to leave this matrix unaltered.

With Q, H, $\hat{X}_0$, and C_0 specified for Monday, the Kalman algorithm was

applied to Monday and then Tuesday. The results are presented in the following section.

Returning to the specification of the covariance matrix H_t, recall that this is associated with the random vector Z_t, which should be interpreted as the noise associated with the observation Y_t. There are at least two reasons why the value $X_t(i)$ of stock i should not necessarily coincide with the transaction price at the same time. The first consideration is the bid-asked spread. It is reasonable to regard the value $X_t(i)$ as falling somewhere between the bid and asked prices. Moreover, if the transaction prices fluctuate between the bid and asked prices, then the transaction prices will fluctuate around the underlying value $X_t(i)$ without necessarily coinciding with it.

A second reason why the stock value and the transaction price need not be the same has to do with ticks. Quoted stock prices are always multiples of some fraction such as 1/8, but it is not reasonable to expect the underlying value of the stock to be a multiple of the same fraction. By this consideration, however, it might be reasonable to assume the transaction price is the multiple that comes the closest to the underlying value.

These two considerations suggest estimates for H_t: we took H_t to be diagonal with all diagonal elements equal to 0.005. We made this choice because the corresponding standard deviation is about 0.07, a number which is consistent with the bid-ask spread and the 1/8 tick size. And note that H_t is positive definite, thereby fulfilling a requirement necessary for implementing the Kalman filtering algorithm.

Actually, since it is difficult to envision a good statistical procedure for estimating H_t, one should probably regard H_t as an exogenous parameter whose value the model builder has some freedom to choose. With

8

this perspective one should understand how the choice of H_t will affect the behavior of the MMI estimates. Based on our experience, it appears H_t should be regarded as analogous to an exponential smoothing parameter, in that its choice affects the sensitivity of the Kalman filtering algorithm. In particular, the smaller the value of the diagonal elements, the more quickly the Kalman estimates will respond to changing prices.

To see this, consider the following simple two stock example. Suppose $\hat{X}_0 = (10, 100)'$,

$$ Q = C_0 = \begin{pmatrix} .01 & .09 \\ .09 & 1.0 \end{pmatrix}, $$

and in period one only the first stock trades, at the price 9.75, which is a 2.5% drop. Note that since the correlation of the stocks is 0.9, one would expect the estimate of the second stock's value to be sharply lower as well. Using the Kalman filtering algorithm, the estimate $\hat{X}_1$ was computed using different values for H_t (which is a scalar in this case), as presented in the following table:

H_t	$X_1(1)$	$X_1(2)$
.5	9.99	99.95
.1	9.98	99.80
.05	9.96	99.63
.01	9.88	98.88
.005	9.83	98.50

Thus with the value 0.005 for H_t (the value we use for the empirical results in the next section), a 2.5% decline in the transaction price of stock 1 leads to 1.7% and 1.5% declines in the estimated values of stocks 1

and 2, respectively.

4. Results

To understand the Kalman estimates of the index, it is instructive to first consider the estimates of the individual stocks (see Figures 3-22, which show the transaction prices and a two standard deviation confidence interval; the Kalman estimates are the mid-points of the confidence interval).

When a stock is trading, the Kalman estimate of the stock's value is almost exactly the transaction price. The least squares estimate of the true value of the stock differs slightly from the transaction price, reflecting reporting errors, the bid-asked spread, and the discreteness of price movements. The difference is affected by the matrix H, which allows the value estimate to be pulled slightly from the observed price, based on the observed transactions in other stocks. This difference is usually too small to be observed on the accompanying graphs. The confidence interval around the estimate reflects reporting errors, the bid-asked spread, and the tick size for the stock.

If a stock does not trade in a particular minute, the technique nevertheless produces a linear least-squares estimate and corresponding (2-sigma) confidence interval for the value of the stock. The estimate will be based on the previous stock price, the movements in the prices of those stocks which do trade, and the estimated covariances of those prices with the missing stock price.

When a stock price is missing for several or many minutes, its estimated value changes over time as the other stock prices move. In addition, the precision of the estimate decreases as the length of time between trades increases. Thus, the confidence intervals, shown on the

10

accompanying individual stock graphs, balloon out when there is a substantial gap in trading. And before a stock opens, its price is estimated using the same filtering technique and the previous day's close (see below).

The computation of the Kalman estimate of the index begins on Monday the 19th and continues through the end of the 20th. As explained in Section 3, individual stock prices are initialized at the Kalman estimate of the Friday close. All the stocks in the index traded in the last few minutes on Friday, so the Kalman estimate of the Monday opening does not substantially differ from the Friday closing prices.

Figures 23 and 24 compare the conventional basis and Kalman basis for Monday and Tuesday, respectively. Figures 25 and 26 show the 2-sigma confidence intervals for the Kalman estimates of the index on Monday and Tuesday, respectively.

Many of the stocks in the index opened late on Monday morning. Until a stock opens, the conventional cash index is computed using the previous day's close. Since those stocks which were trading showed declines on Monday, the conventional cash index was probably an overestimate of the true value of the underlying stocks. This is one of the factors reflected in the large negative cash-futures spread or basis observed on that day (see Figures 1 and 23).

For the Kalman filter estimate of the index, missing stock prices are not assumed to have remained constant (this is implicitly assumed when using a last-trade based estimate of cash values). Thus, Kalman estimates of the index for Monday morning are lower than the conventional cash index. The negative basis is substantially reduced on Monday morning (see Figure 23). After all the MMI stocks opened on the 19th there were only brief gaps in trading for most stocks, so after the opening the Kalman estimate of the

11

index is not much different from the conventional cash index, and the confidence intervals are small.

No allowance is made in these calculations for the time lapse from close to open. The previous Friday's close is used as an initial value for the individual stocks, and the weekend gap in trading is treated as though it were only a minute long. This results in an unrealistically narrow confidence interval for the Kalman index estimate at the start of trading on both Monday and Tuesday. This problem will be corrected in future work.

On Tuesday morning there was a brief rally in stock prices. Since several stocks opened late, the conventional cash index is probably understated, and the Kalman estimate is above it. Thus, relative to the value of the futures, the Kalman spread is smaller than the conventional cash spread at the open. Then, as futures prices fall, the Kalman filter index is slow to respond, and so the lower conventional cash index leads to the conventional spread being closer to zero for a brief period. For the rest of the day the Kalman spread is slightly closer to zero, or about the same distance from zero, as the conventional spread.

For several of the most important stocks in the index, there are substantial post-opening trading gaps on Tuesday. For example, Dow Chemical stops trading at 10:43 with a transaction at $65.50, and it does not resume trading until 12:32 with a transaction at $68.50. Using linear interpolation would lead one to conclude that the value of Dow was almost constant during this two hour period. The Kalman estimates, by contrast, take into account what is known of the behavior of other prices at the time. During the two hour gap in trading, the estimated value of Dow Chemical drops, then increases, and then drops again. Although the magnitudes of these fluctuations are not as great, note the similarity to the behavior of the futures price over the same period.

During this large gap in trading, the confidence interval widens to more than 10% of the estimated value of Dow. The other three most heavily weighted stocks, IBM, Phillip Morris, and Merck, also have substantial trading gaps during the morning of the 20th. These four stocks make up more than a third of the index value.

This degree of uncertainty about the value of a substantial part of the index is reflected in the confidence intervals for the index itself (see graphs 25 and 26). During Tuesday morning when there were large gaps in trading for some of the major stocks, the confidence interval for the estimate of the index itself is large.

5. Concluding Remarks

The Kalman filtering method described in this paper works reasonably well for estimating the value of a cash index when there are gaps in trading, at least when judged by comparing the conventional cash-futures spread against the Kalman spread. The Kalman spread is much closer to zero during the Monday opening, and Kalman filtering outperforms the conventional, last-trade method for most of Tuesday. Additional experimentation with the model and parameter estimates will be done in a further study of the Tuesday estimates.

An important contribution of the Kalman filtering method is that it provides a good measure of the degree of uncertainty about the index value (as shown in Figures 25 and 26). The 2-sigma confidence intervals used throughout our paper could, of course, be replaced by other interval estimates. Such intervals could be useful for a variety of purposes, including arbitrage trading. In particular, one might judge whether there is a legitimate arbitrage opportunity by comparing not only the futures price with an estimate of cash value, but, in addition, looking to see if

13

the futures price falls within the confidence interval.

The confidence intervals for October 19 and 20, 1987, may also provide some insight into events during the stock market crash. Since the futures price was often outside the estimated confidence interval, one might conclude that nonsynchronous trading was not an important factor in explaining the large spreads.

One potential elaboration of the Kalman technique would involve correction for selection bias. Selection bias correction (see Heckman, 1979) is useful in cases where missing data are suspected of being systematically lower or higher than average. For example, the true values of stocks for which transaction prices are missing during the trading hiatus on the morning of the 20th are probably systematically more depressed than those stocks which continued to trade. We can identify a systematic bias in this case because we know from the government reports that trading in many stocks was shut down because of an excess of sell orders over buy orders. The Kalman technique assumes that missing values bear their normal relation to other data; it does not allow for a bias of this kind.

The Kalman filtering technique can be applied to the calculation of index values in other situations where non-synchronous trading is significant. Non-synchronous trading would not normally be a significant factor in pricing the MMI futures, since the stocks in the index are relatively liquid. However, pricing the S&P 500 cash index before all 500 stocks have opened (perhaps for the purpose of establishing a cash-futures arbitrage position) is a straight-forward application of the same technique. A slight modification of this technique would employ early morning London prices to estimate the movement of NYSE stocks before they open.

14

Perhaps a more important application will arise when futures contracts on international indices begin trading. Estimating the true value of, for instance, the Capital International EAFE index involves formidable missing data problems. Many of the European equity markets are thin by U.S. standards. In addition, because the index covers trading in time zones from London to Australia, the stock prices which make up the index are frequently very stale. Further, national holidays vary from country to country. The Kalman filtering technique would enable a potential arbitrageur to estimate the value of the index. It would also permit the exchange to post index values when components of the index are not trading.

References

Bassett, Jr., G. W., V. G. France, and S. R. Pliska (1989). The MMI Cash-Futures Spread on October 19, 1987. Working Paper No. 16, Office for the Study of Futures and Options, University of Illinois at Chicago. To appear in *The Review of Futures Markets*.

Davis, M. H. A., *Linear Estimation and Stochastic Control*, Academic Press, New York, 1977.

Edwards, F. R., (1988). Studies of the 1987 Stock Market Crash: Review and Appraisal. Working Paper No. 168, Center for the Study of Futures Markets, Columbia University.

Heckman, J. J., Sample Selection Bias as a Specification Error, *Econometrica* 47 (1979), 153-161.

Krishnan, V., *Nonlinear Filtering and Smoothing: An Introduction to Martingales, Stochastic Integrals, and Estimation,* John Wiley and Sons, New York, 1984.

Stock	Symbol	σ
Merck & Co.	MRK	.4234
International Business Machines	IBM	.2053
Phillip Morris	MO	.1702
Dow Chemical	DOW	.2826
Procter & Gamble	PG	.1946
Du Pont	DD	.2423
Johnson & Johnson	JNJ	.1773
Minnesota Mining & Manufacturing	MMM	.1922
General Motors	GM	.1664
Eastman Kodak	EK	.2084
General Electric	GE	.1358
International Paper	IP	.1253
Exxon	XON	.0918
Coca Cola	KO	.1311
Chevron	CHV	.0902
Mobil	MOB	.1132
Sears Roebuck	S	.0889
USX Corporation	X	.0687
AT&T	T	.1007
American Express	AXP	.1122

Table 1   Estimated standard deviations of one minute price
differences for MMI stocks.

Figure 1   Futures and Cash Indices

October 19th, 1987
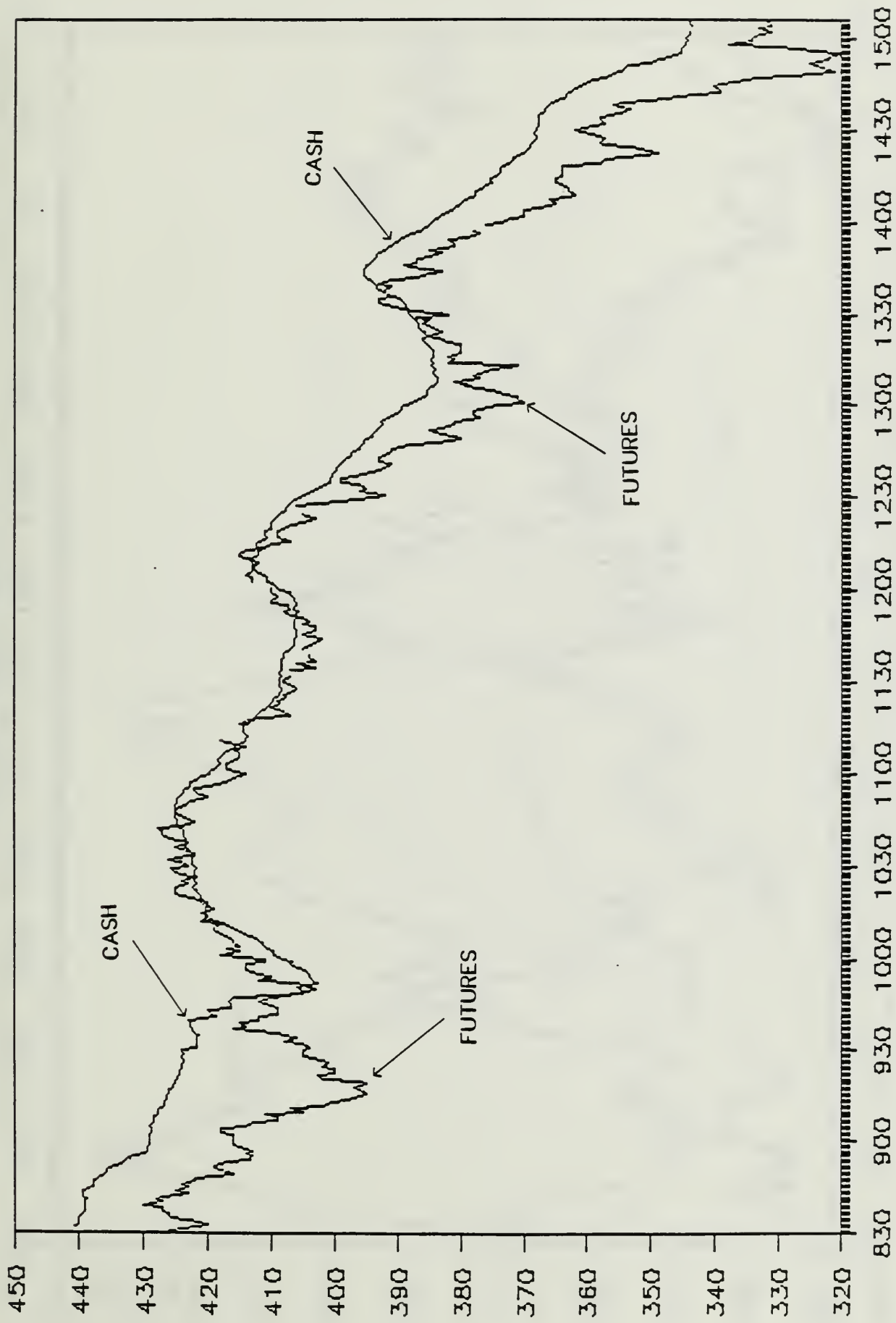
Figure 2 Futures and Cash Indices

October 20th, 1987

Figure 3

AXP on October 20th, 1987
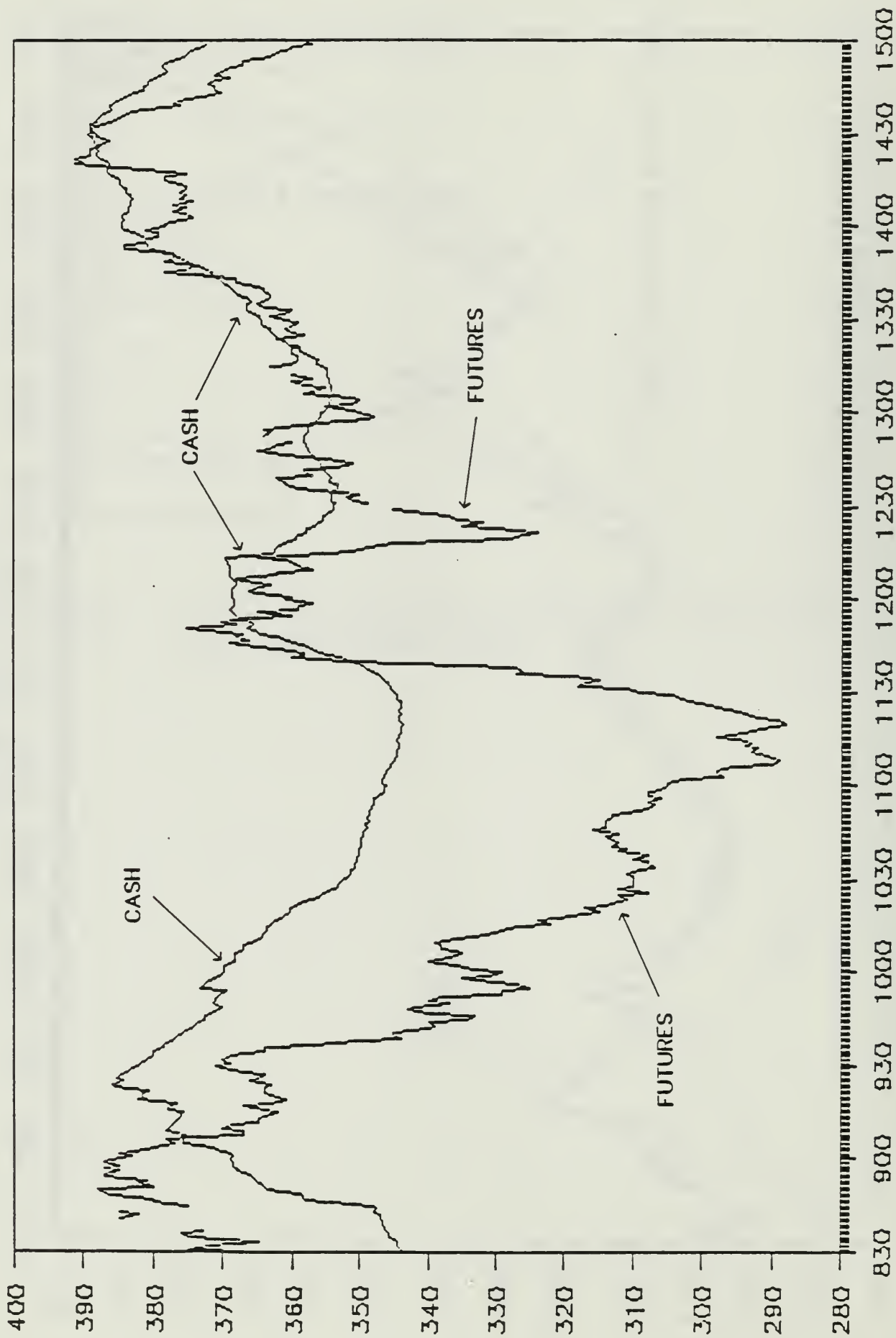
2-Sigma Confidence Intervals
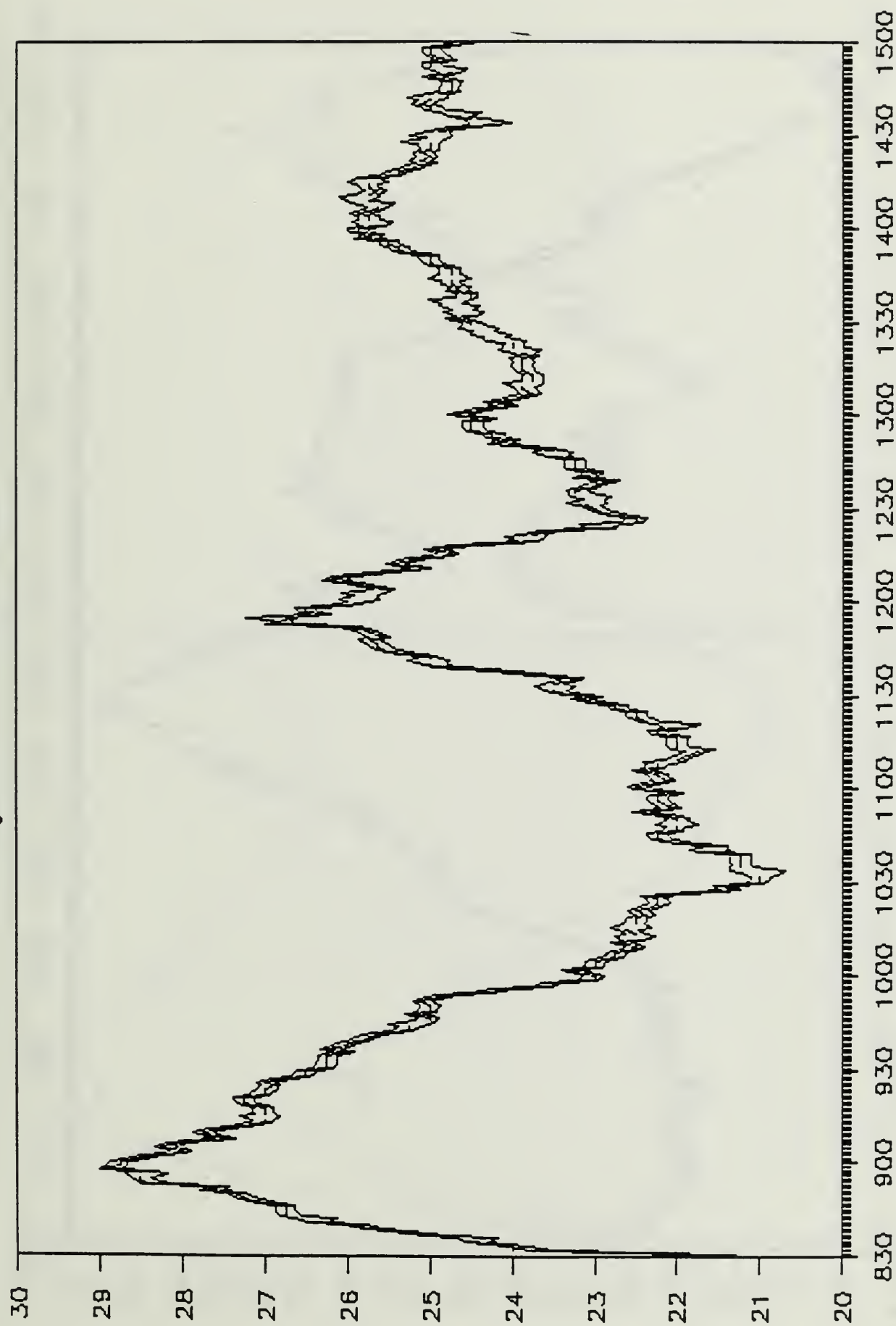
Figure 4   CHV on October 20th, 1987

2-Sigma Confidence Intervals
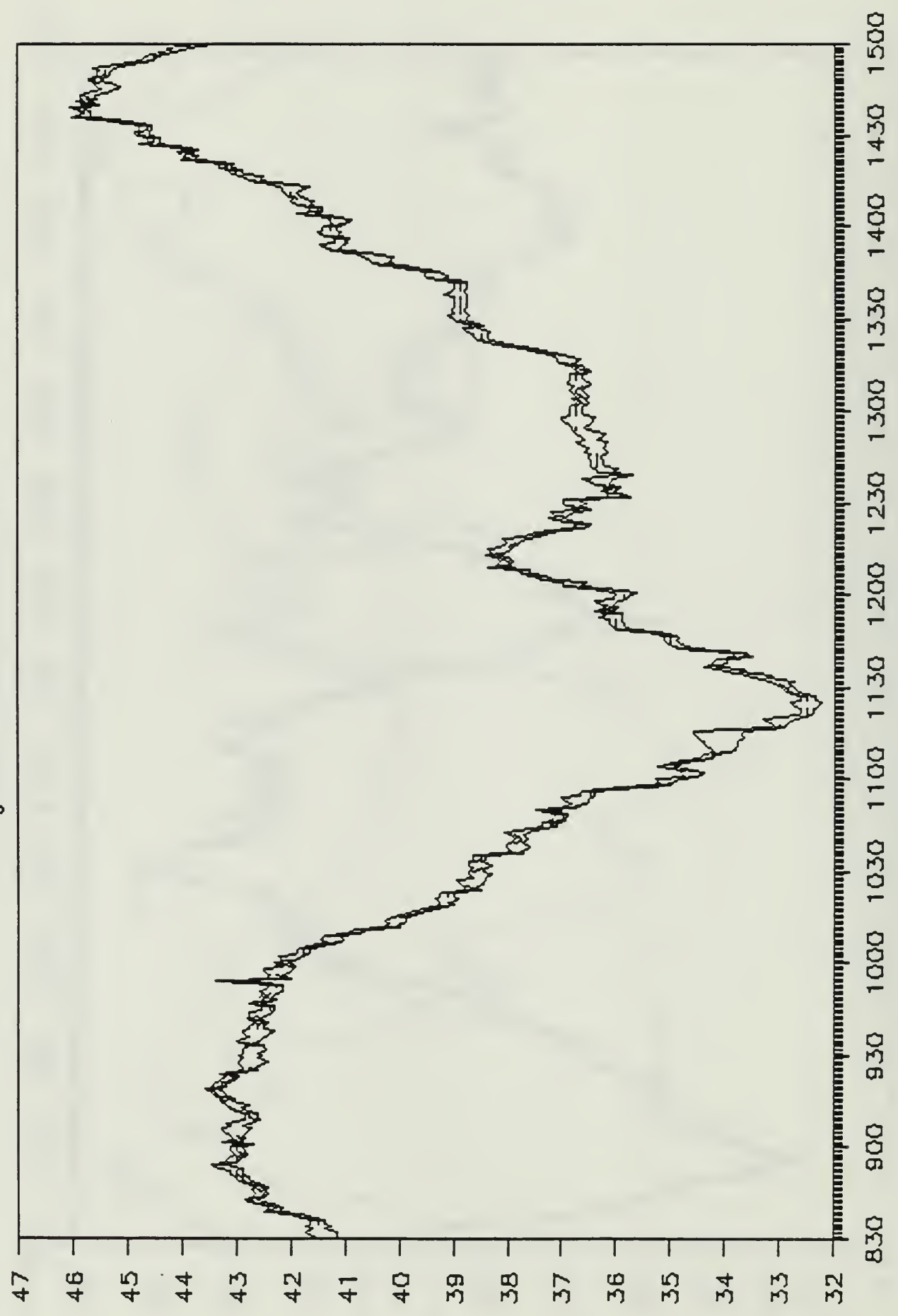
Figure 5 DD on October20th, 1987
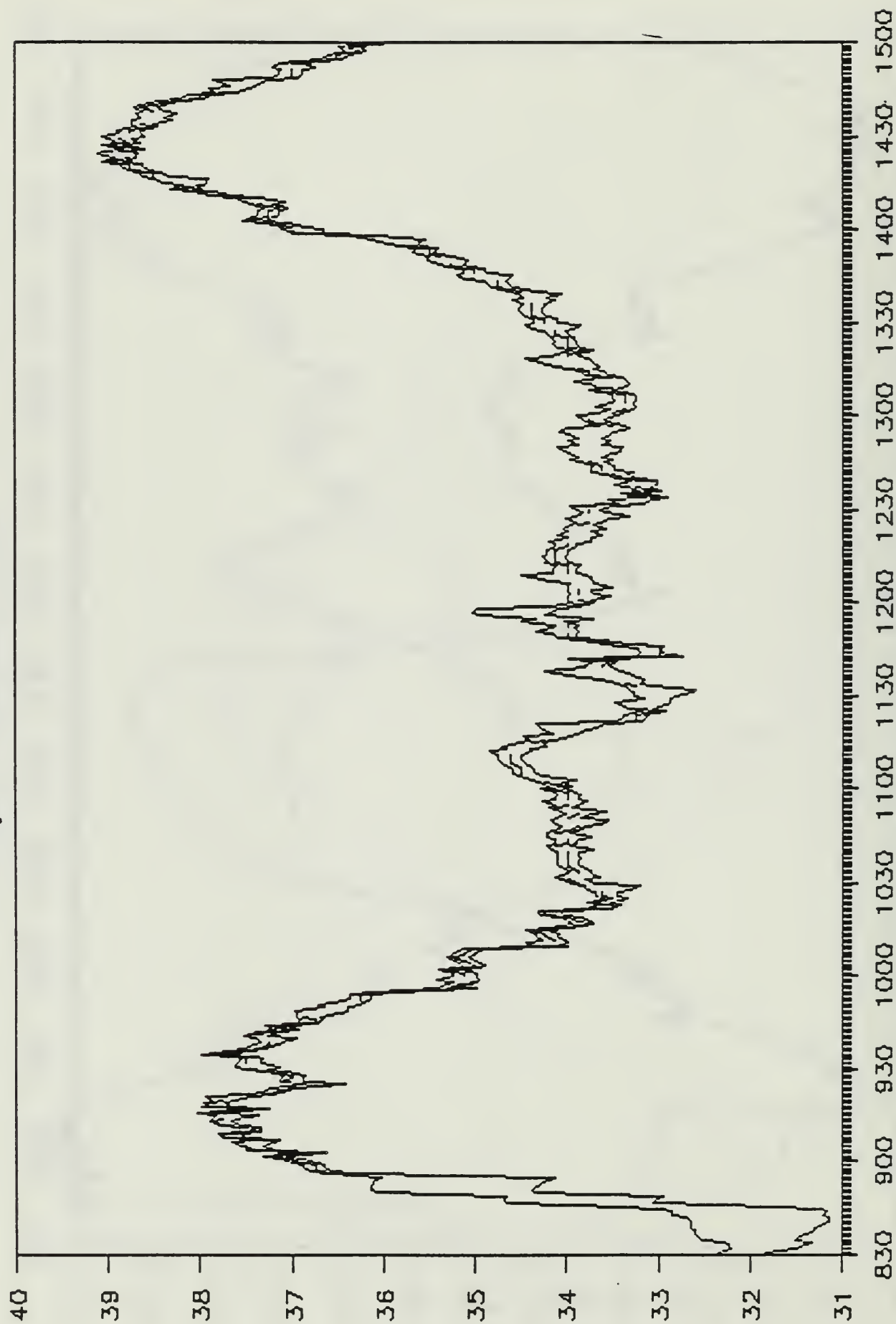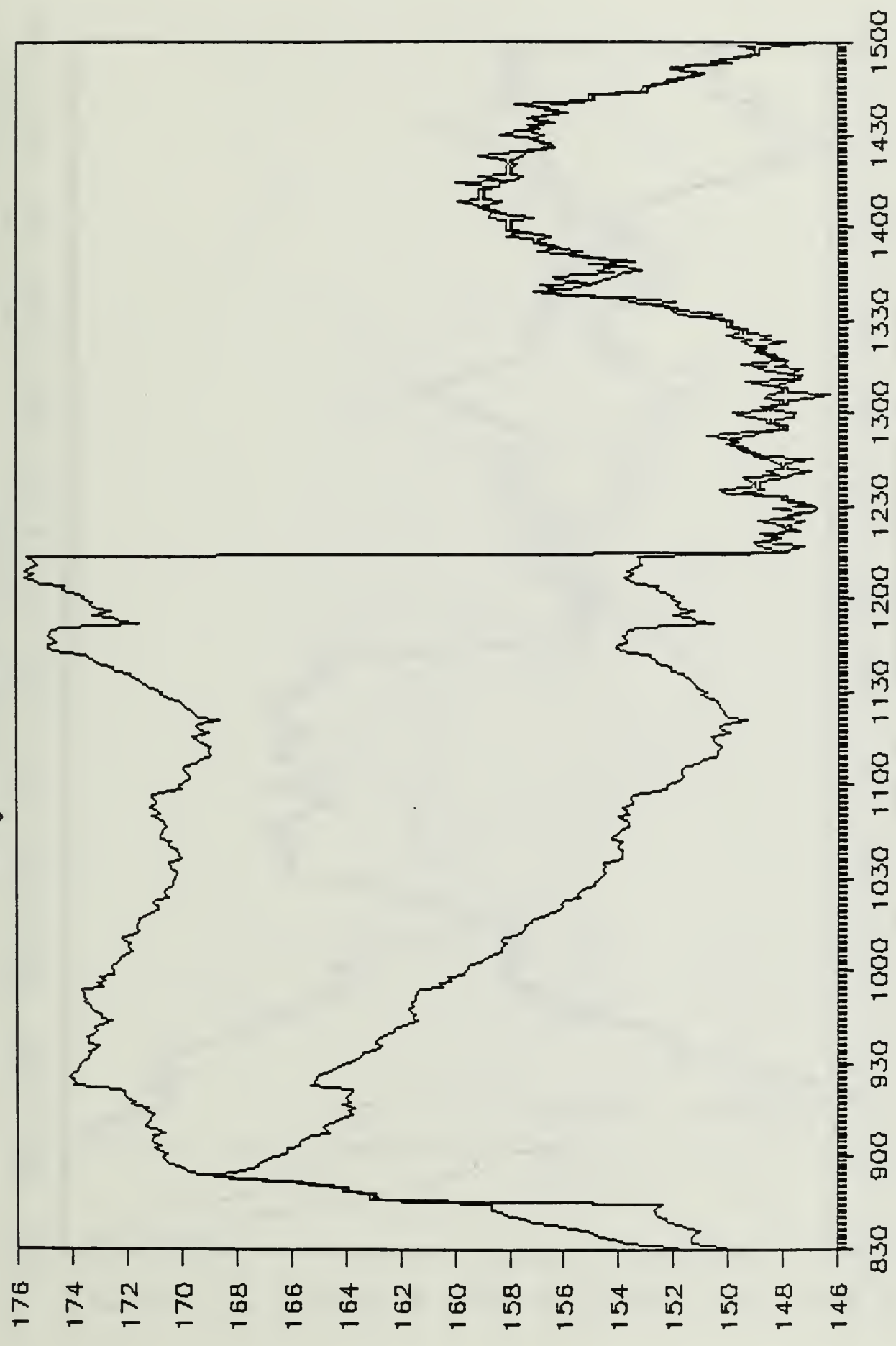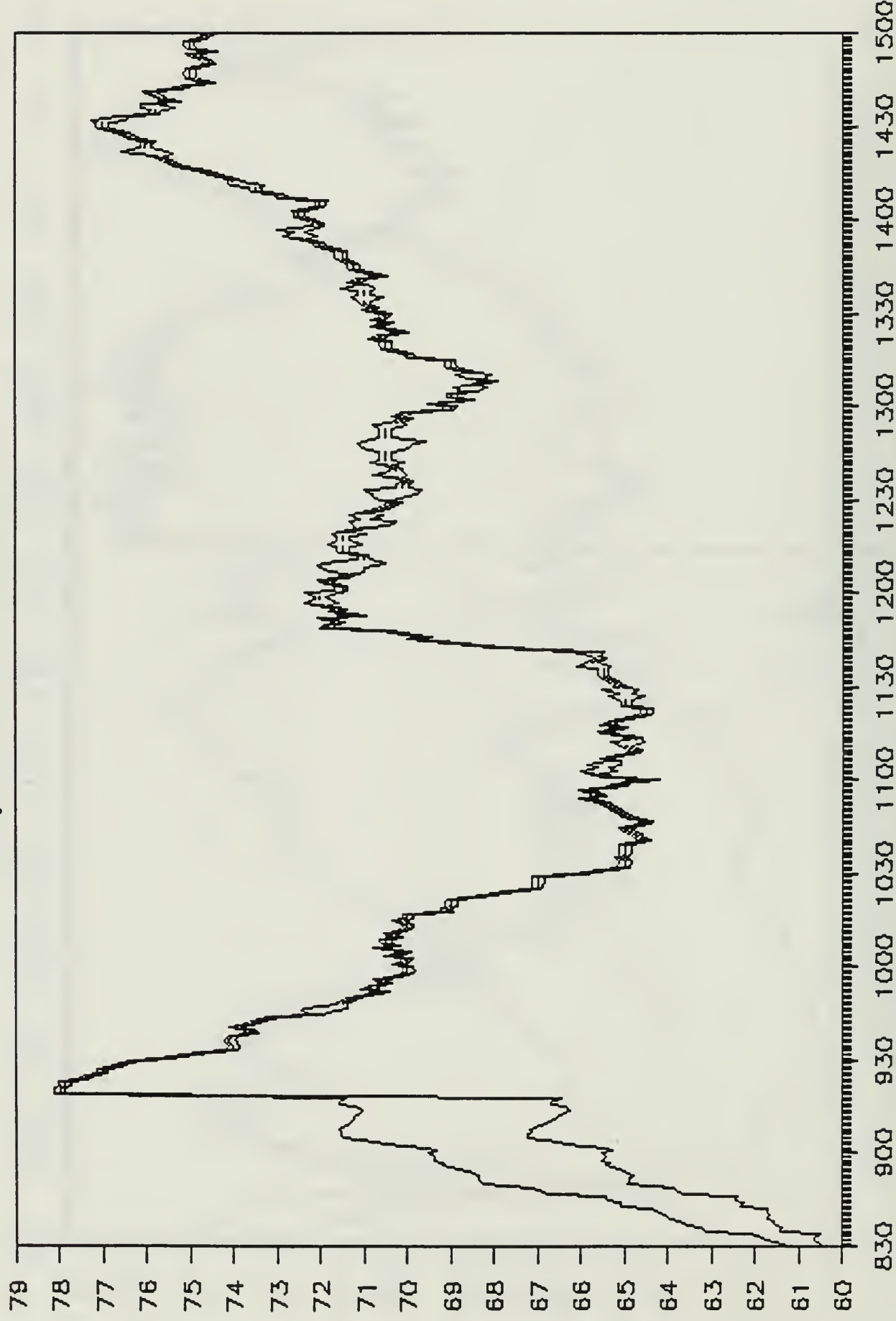
2—Sigma Confide:e Intervals
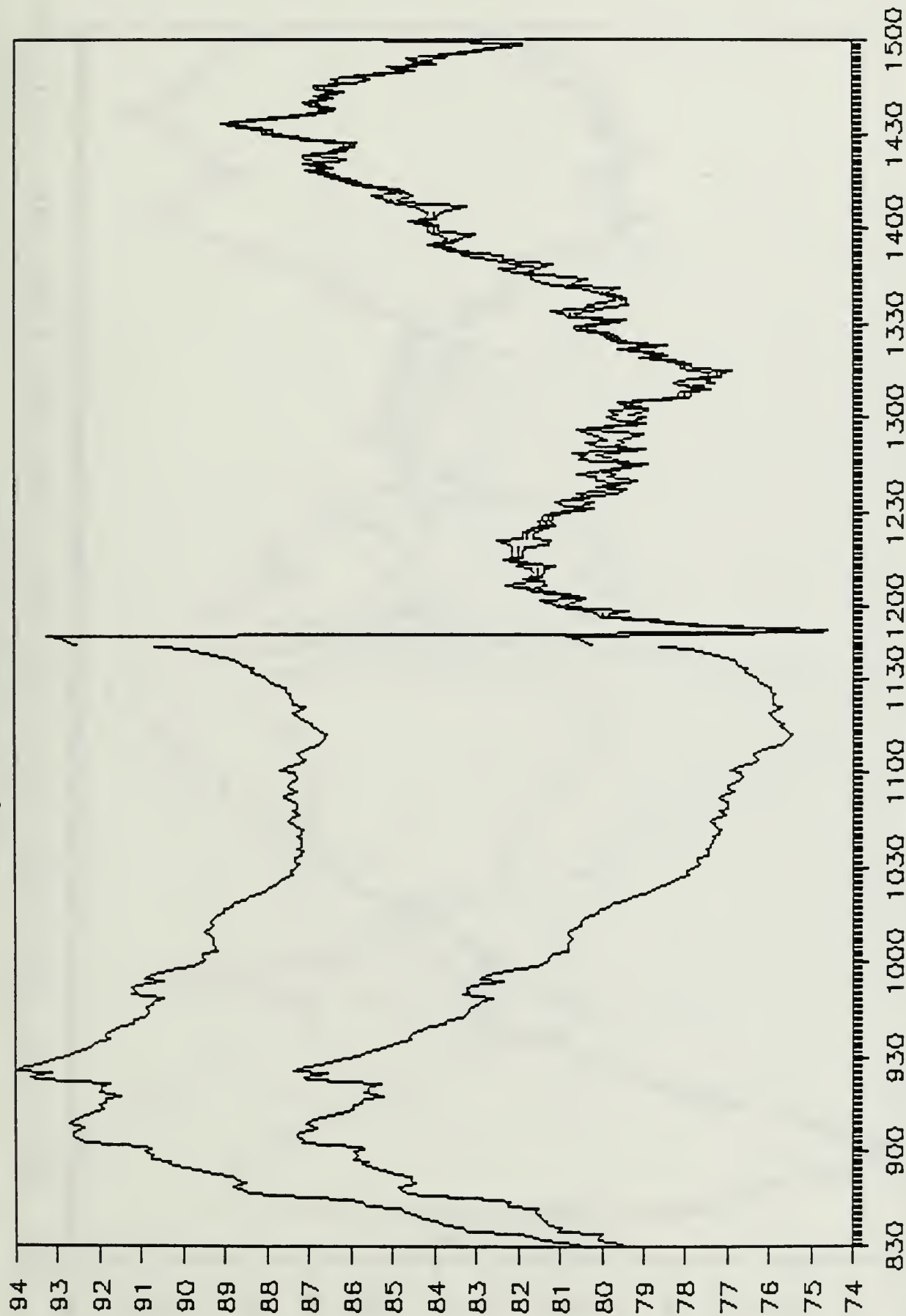
Figure 6   DOW on October 20th, 1987

2-Sigma Confidence Intervals
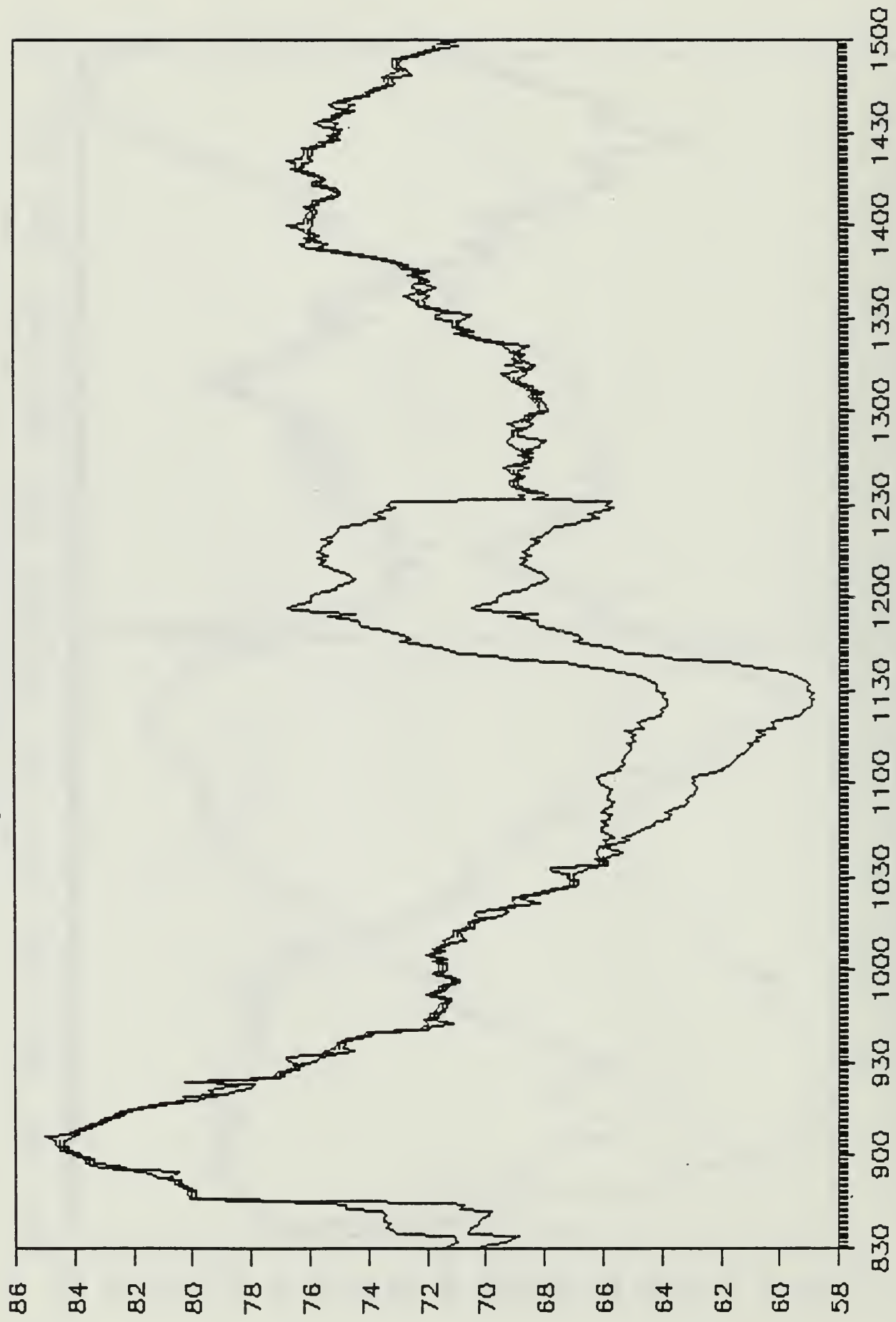
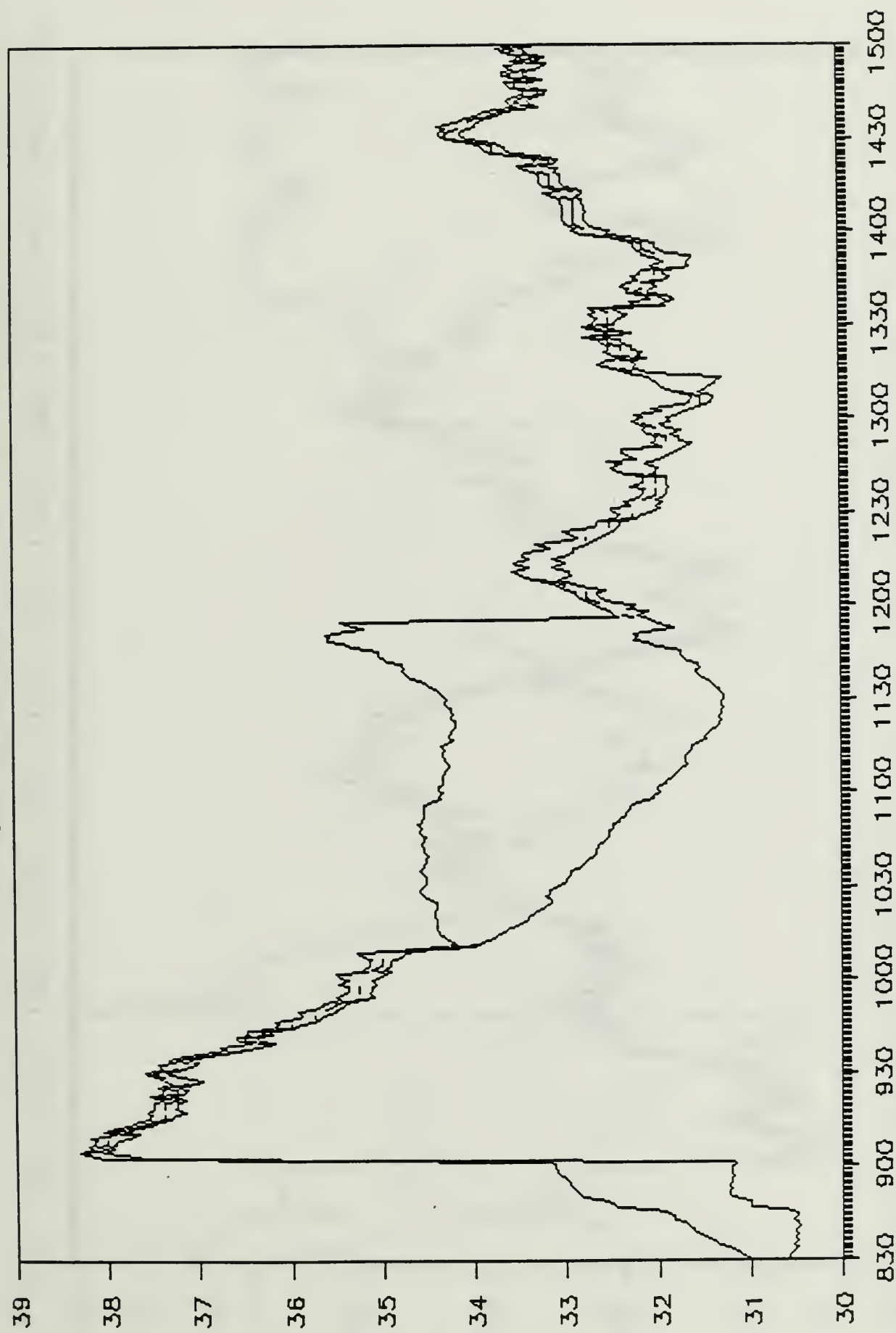Figure 7   EK on October 20th, 1987

2-Sigma Confidence Intervals
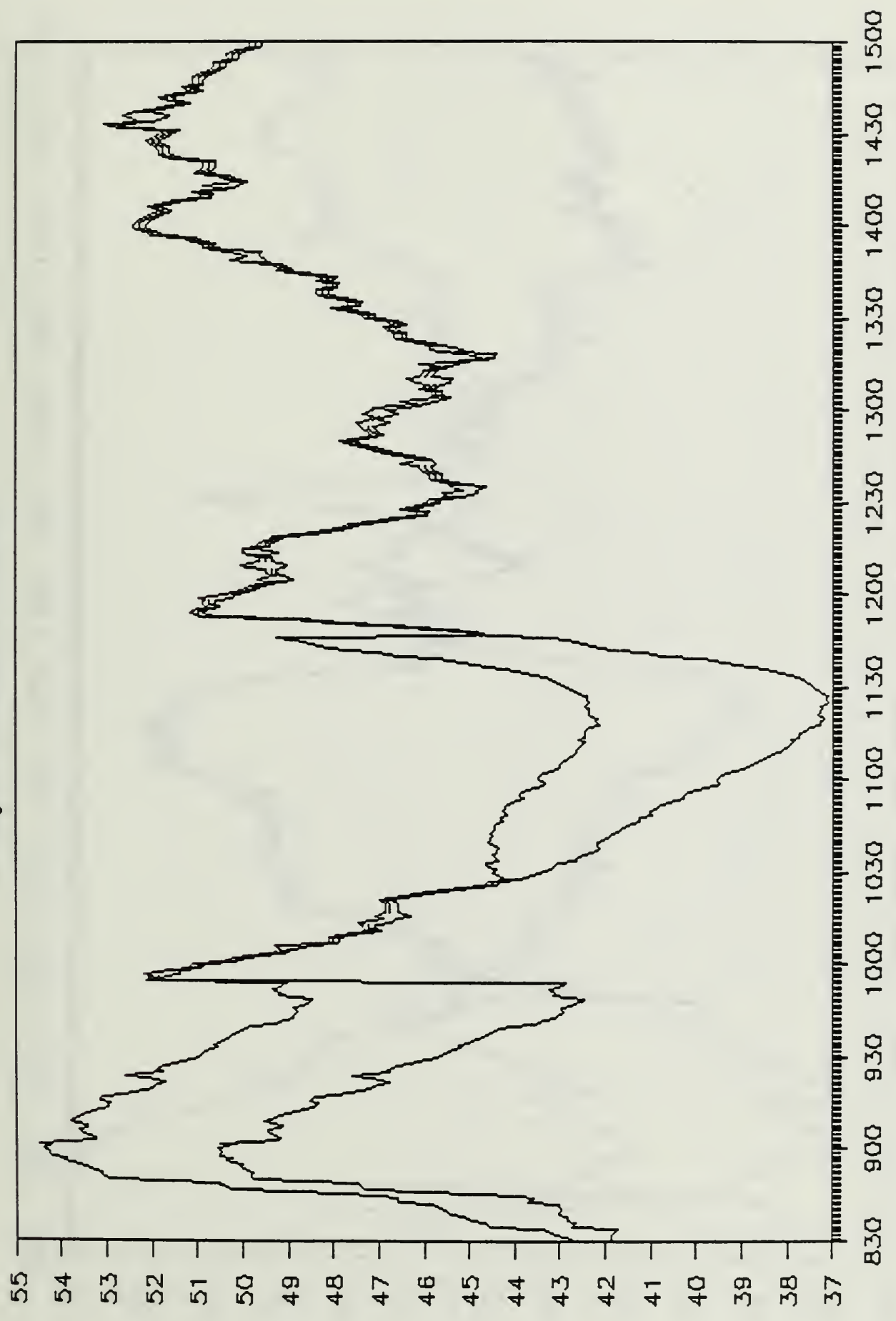
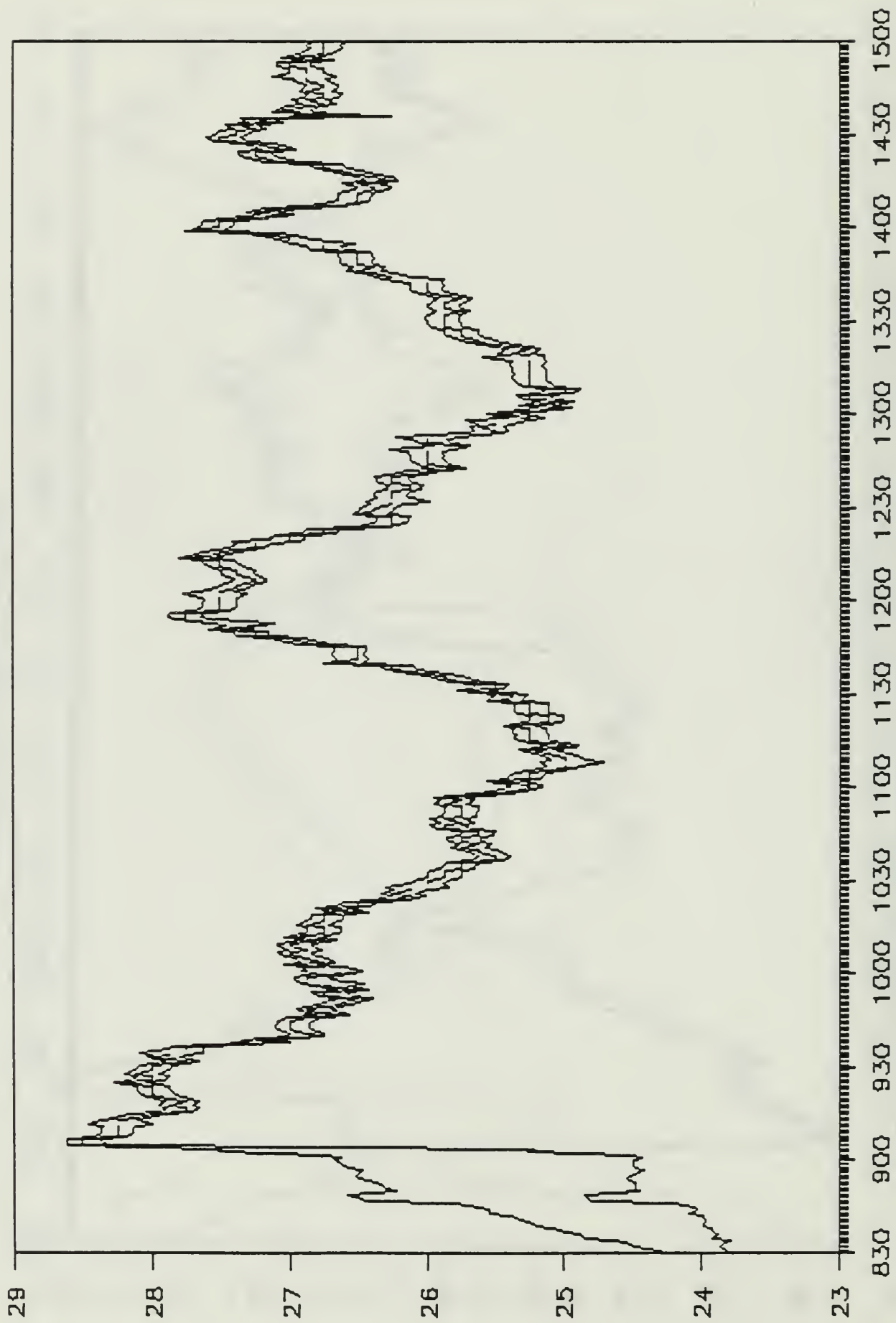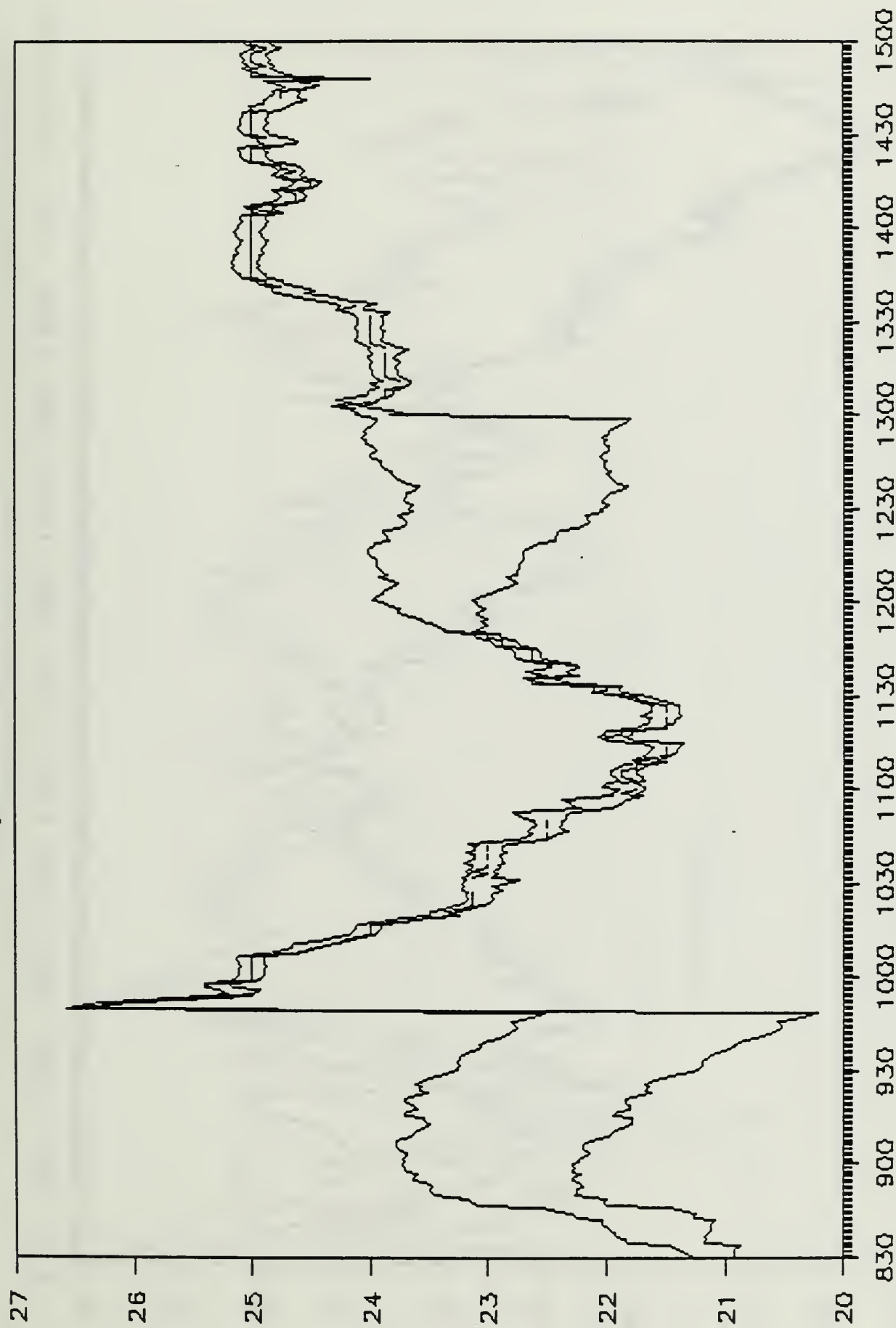Figure 8   GE on October 20th, 1987

2-Sigma Confidence Intervals
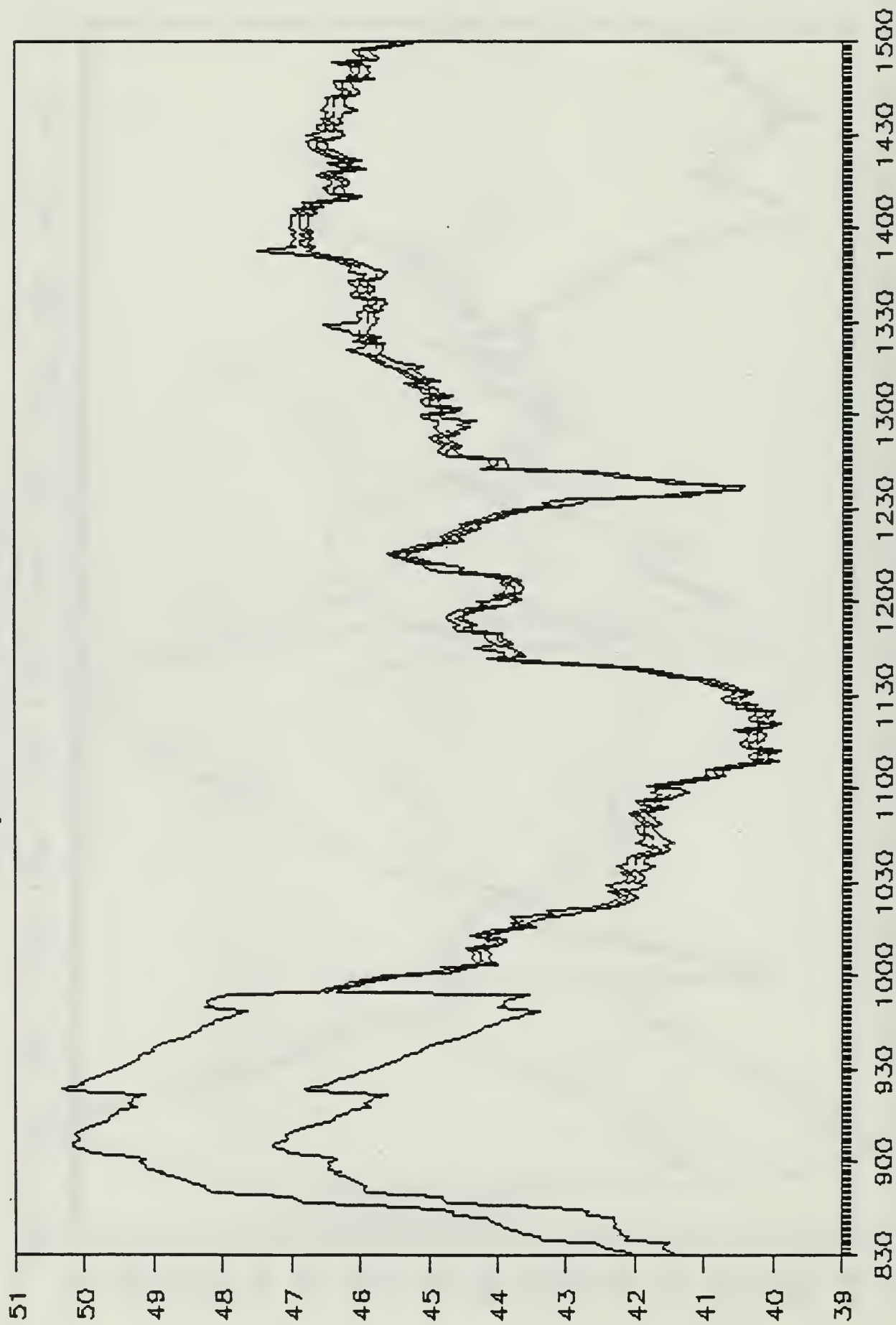
Figure 9   GM on October 20th, 1987

2-Sigma Confidence Intervals
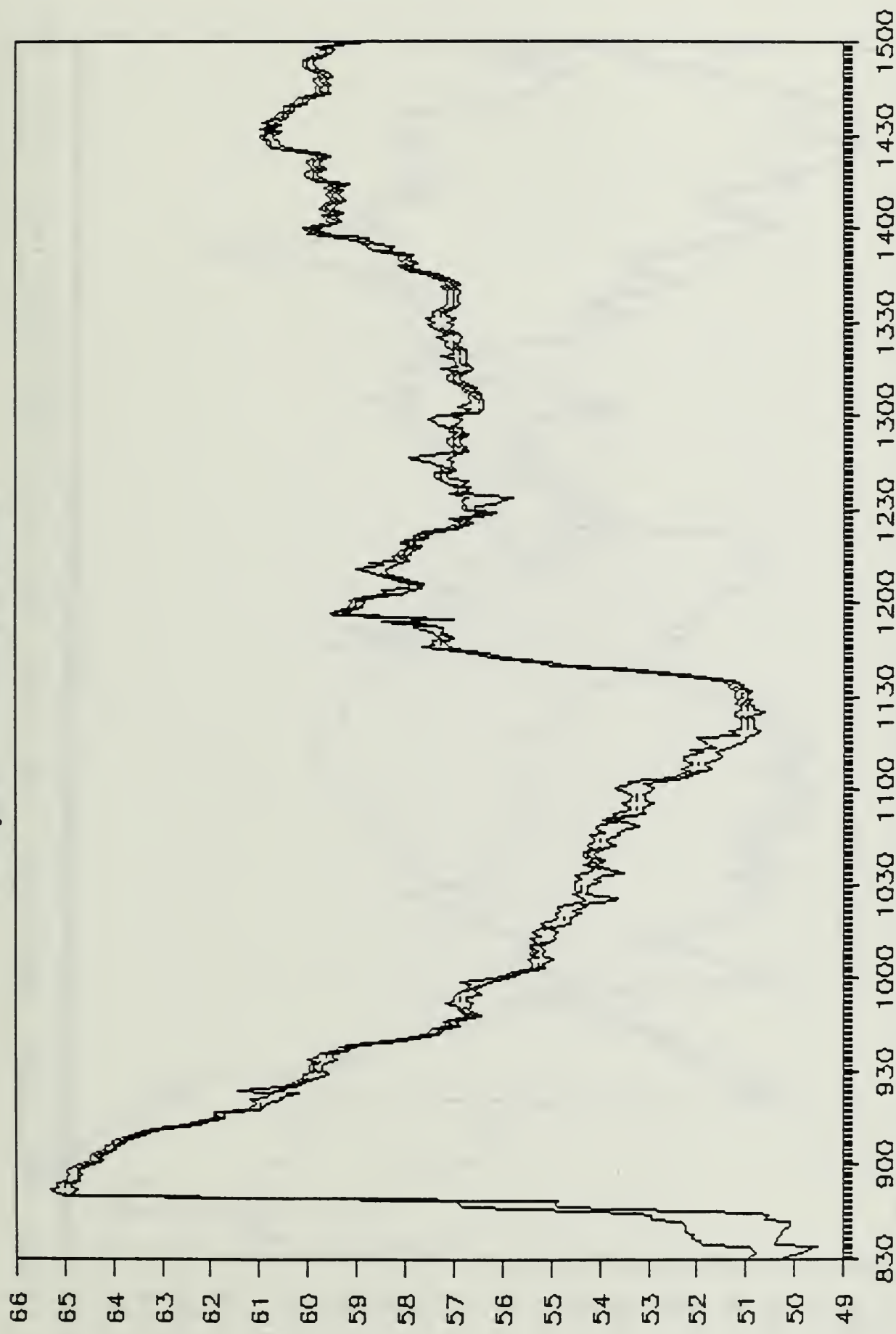
Figure 10   IBM on October 20th, 1987

2-Sigma Confidence Intervals
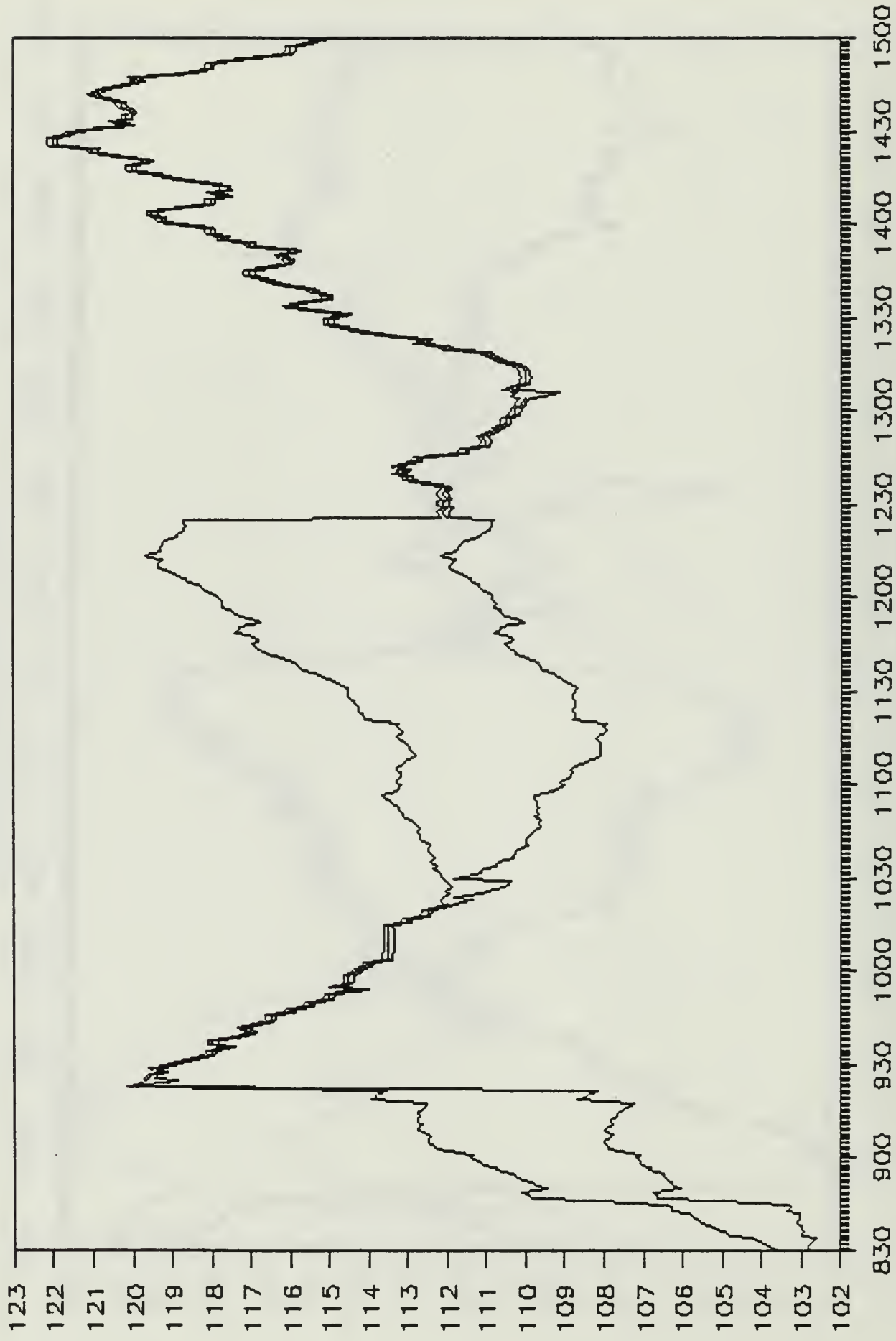
Figure 11   IP on October 20th, 1987

2–Sigma Confidence Intervals
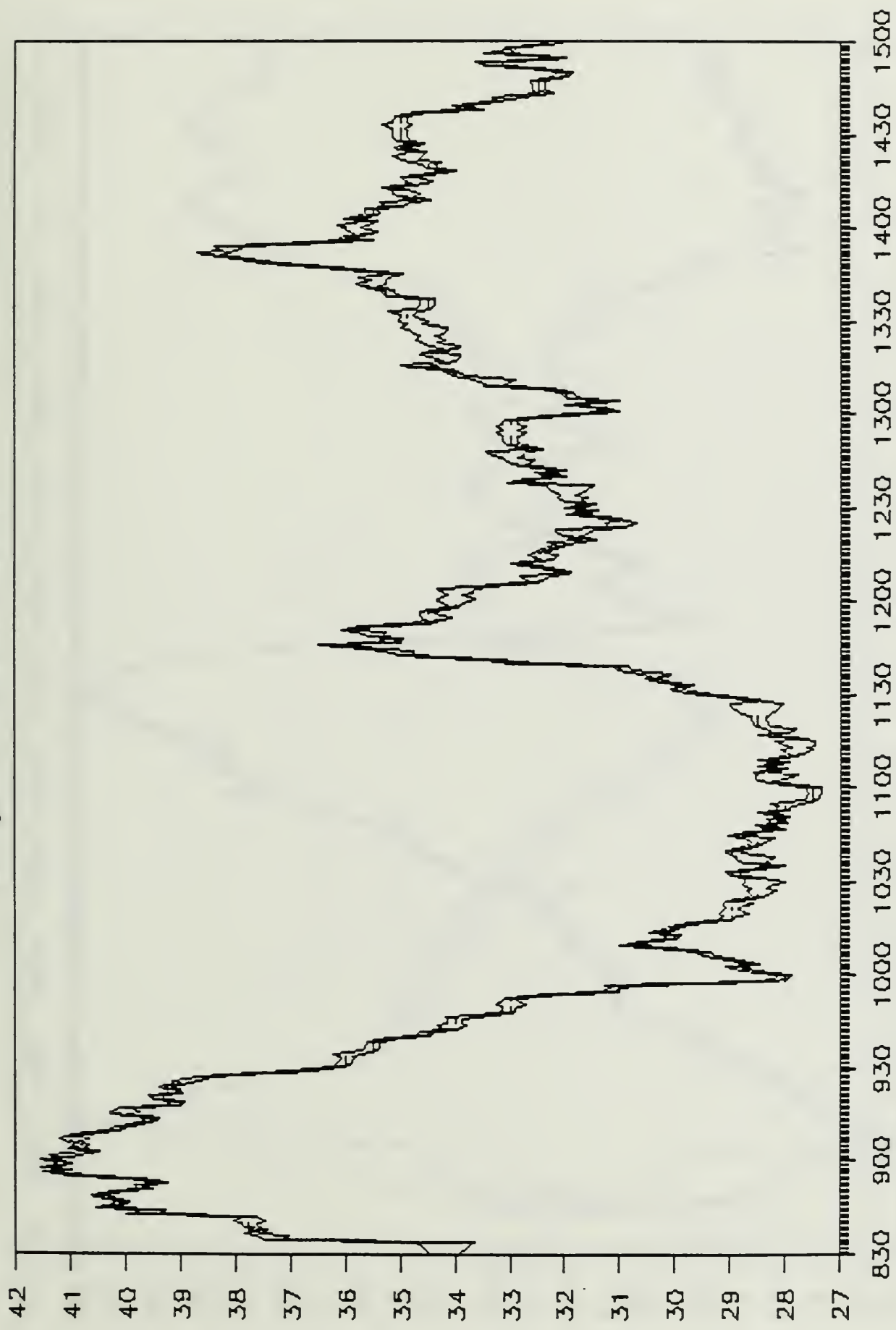
Figure 12   JNJ on October 20th, 1987

2-Sigma Confidence Intervals
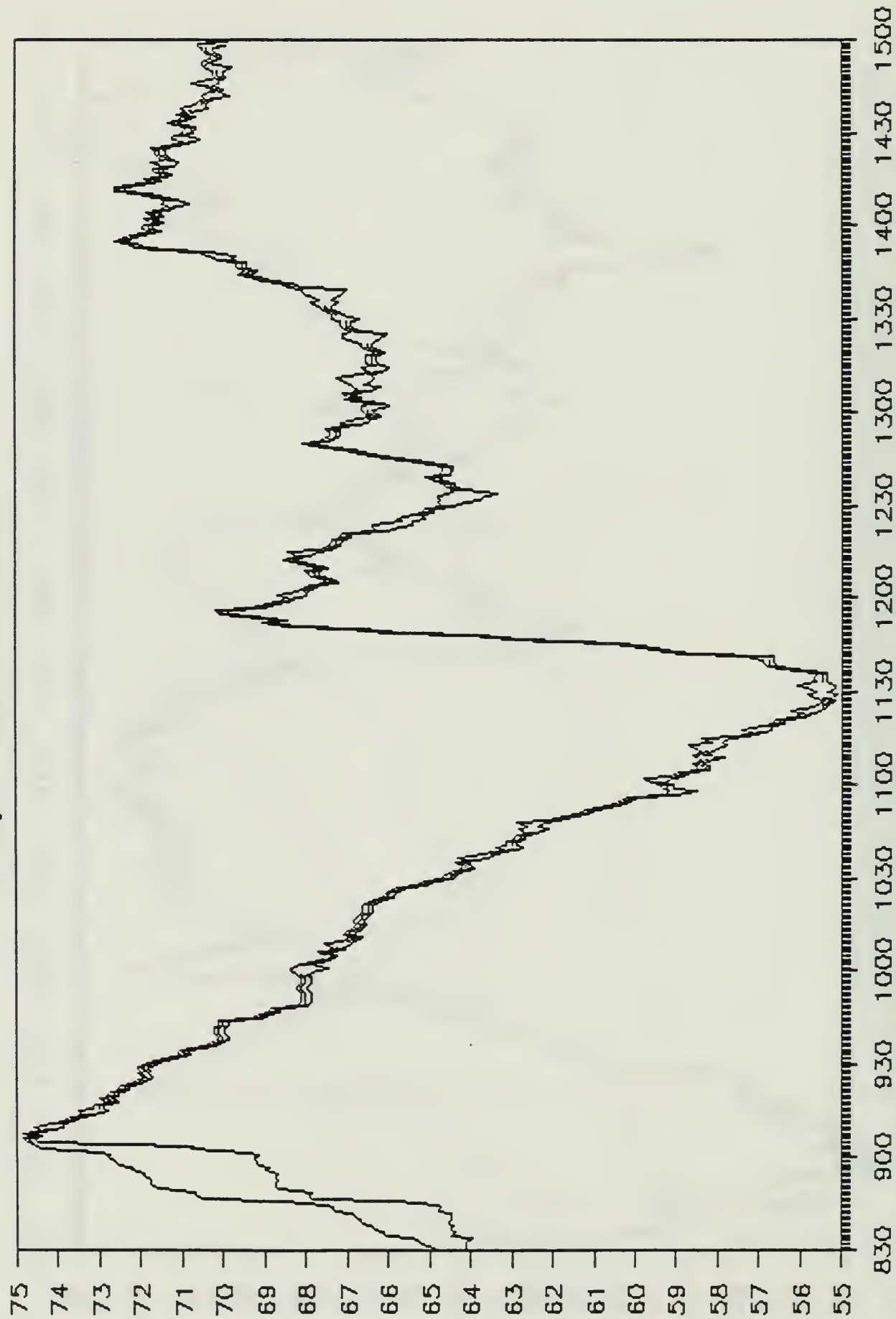
Figure 13   KO on October 20th, 1987

2—Sigma Confidence Intervals
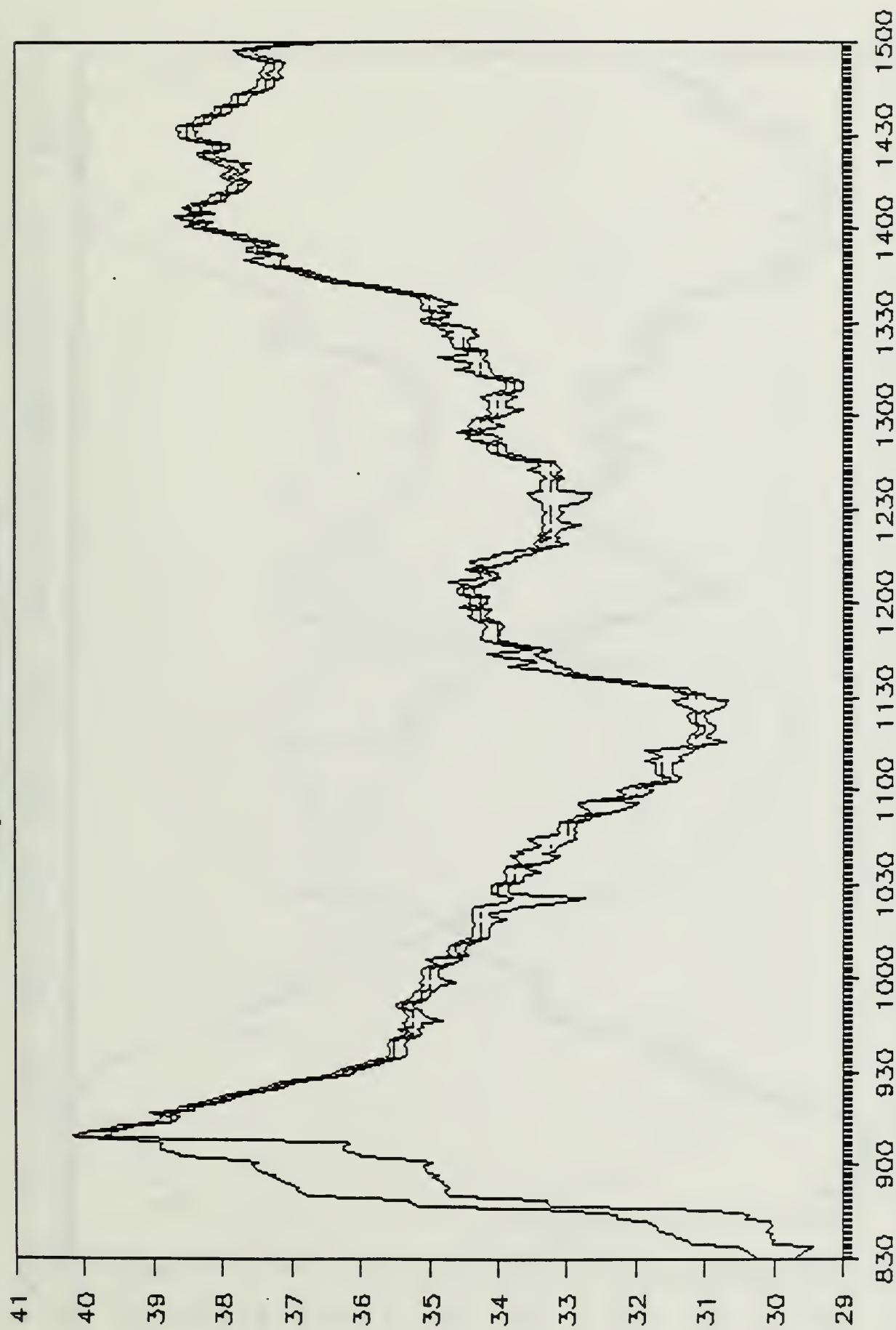
Figure 14

MMM on October 20th, 1987

2-Sigma Confidence Intervals
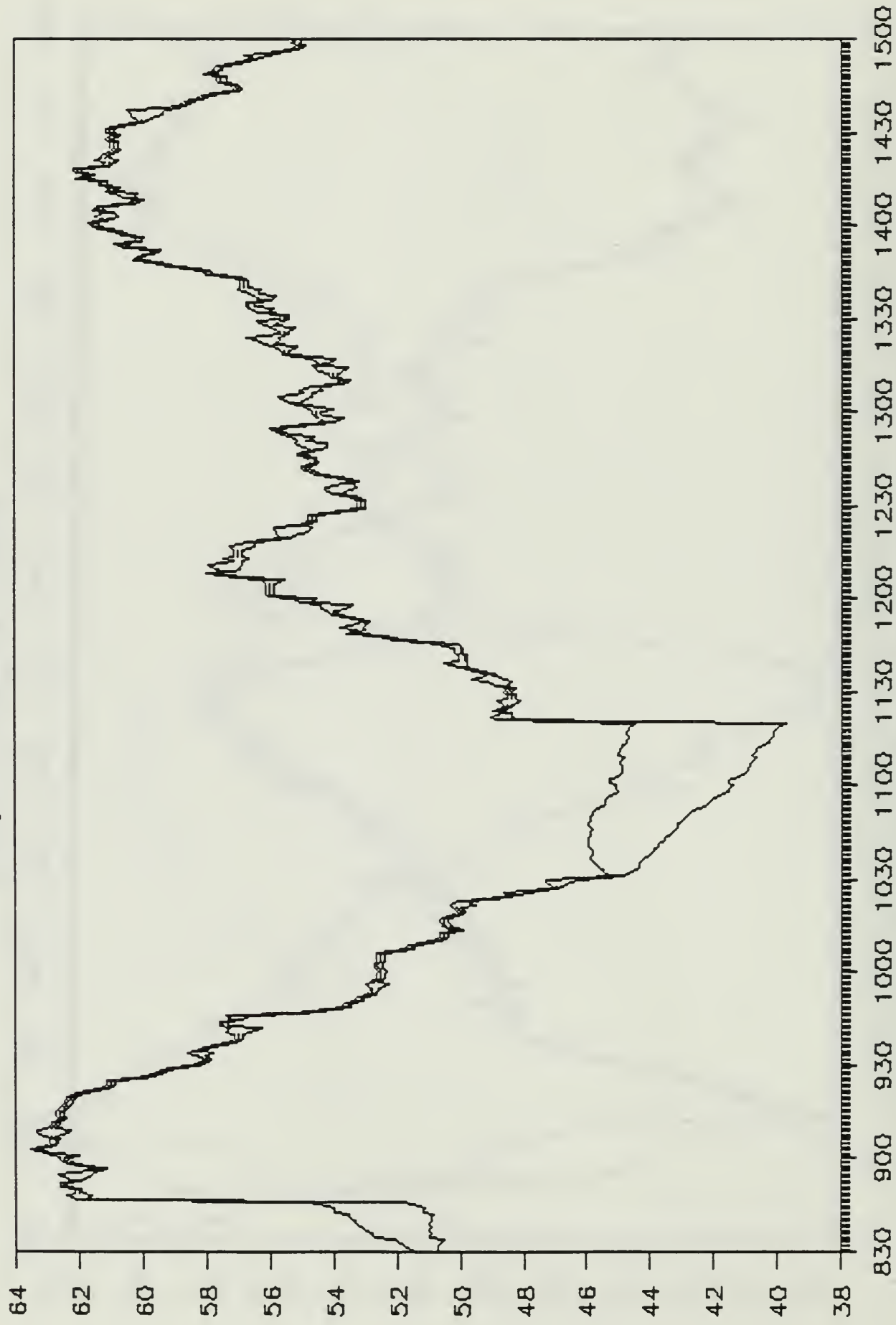
Figure 15   MO on October 20th, 1987

2-Sigma Confidence Intervals
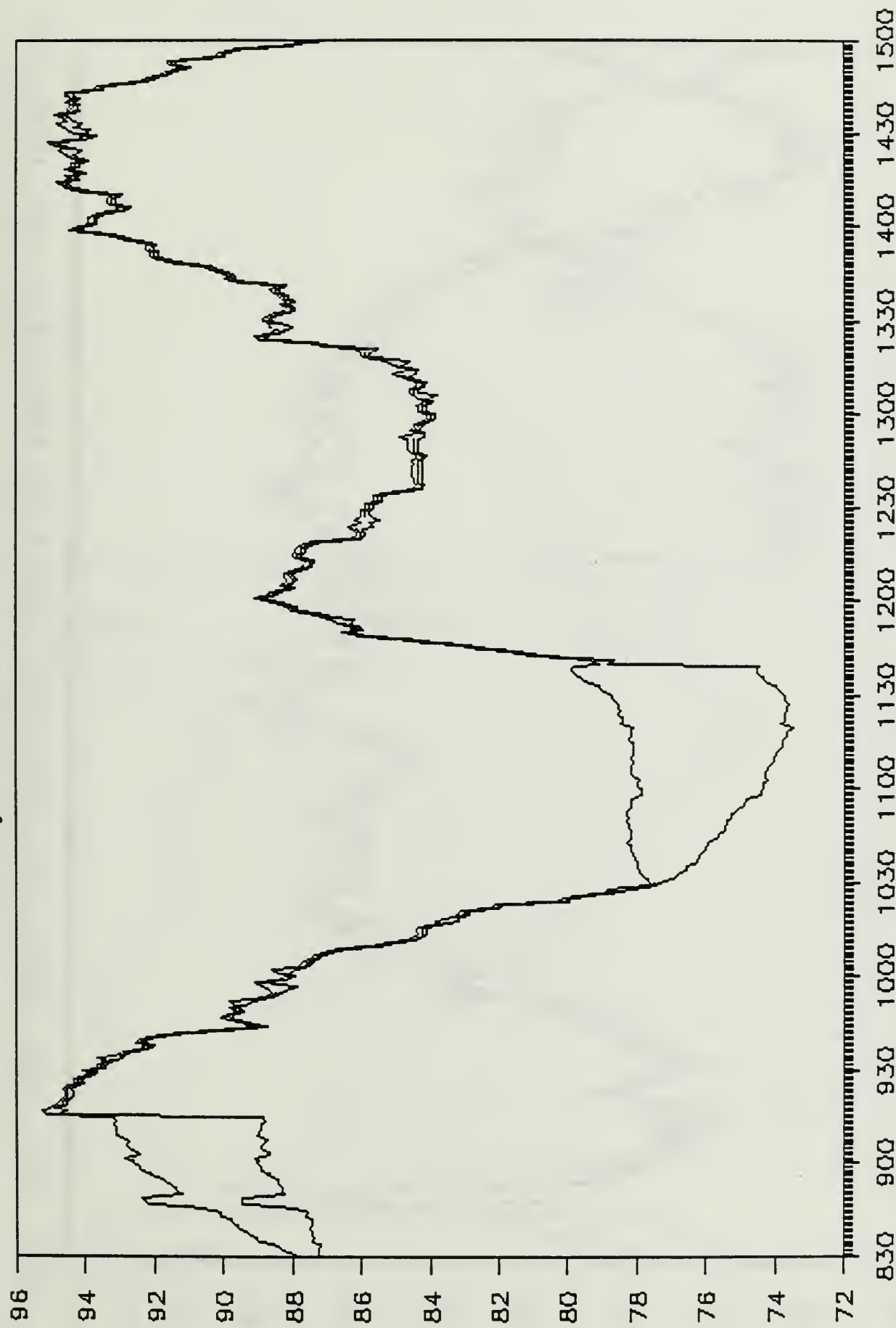
Figure 16

MOB on October 20th, 1987

2-Sigma Confidence Intervals

Figure 17   MRK on October 20th, 1987

2-Sigma Confidence Intervals

Figure 18   PG on October 20th, 1987

2-Sigma Confidence Intervals

Figure 19   S on October 20th, 1987

2-Sigma Confidence Intervals

Figure 20   T on October 20th, 1987

2-Sigma Confidence Intervals

Figure 21   X on October 20th, 1987

2—Sigma Confidence Intervals

Figure 22   XON on October 20th, 1987

2-Sigma Confidence Intervals
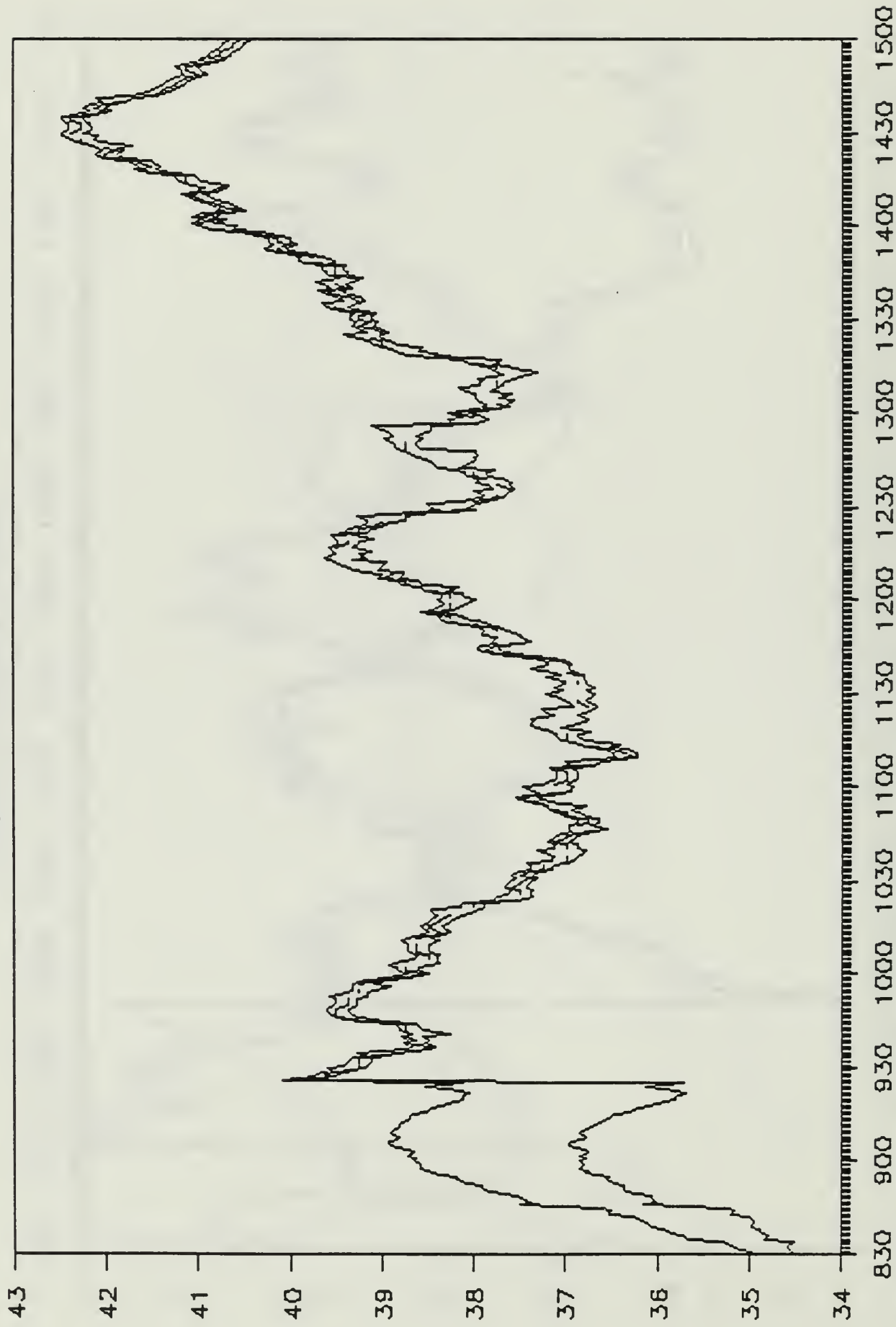
Figure 23

Cash— and Kalman—Futures Basis

October 19th, 1987
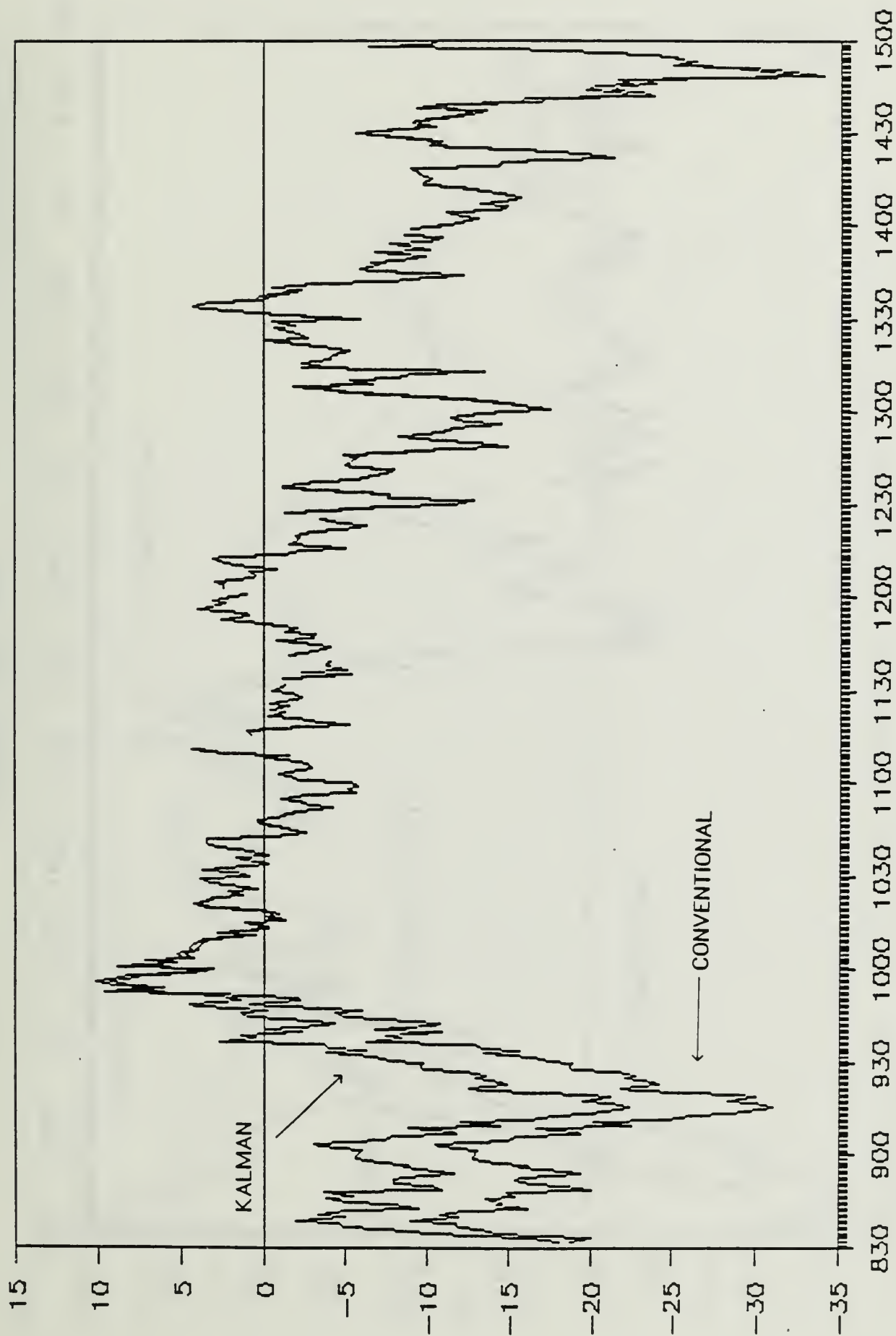
Figure 24

Cash– and Kalman–Futures Basis

October 20th, 1987
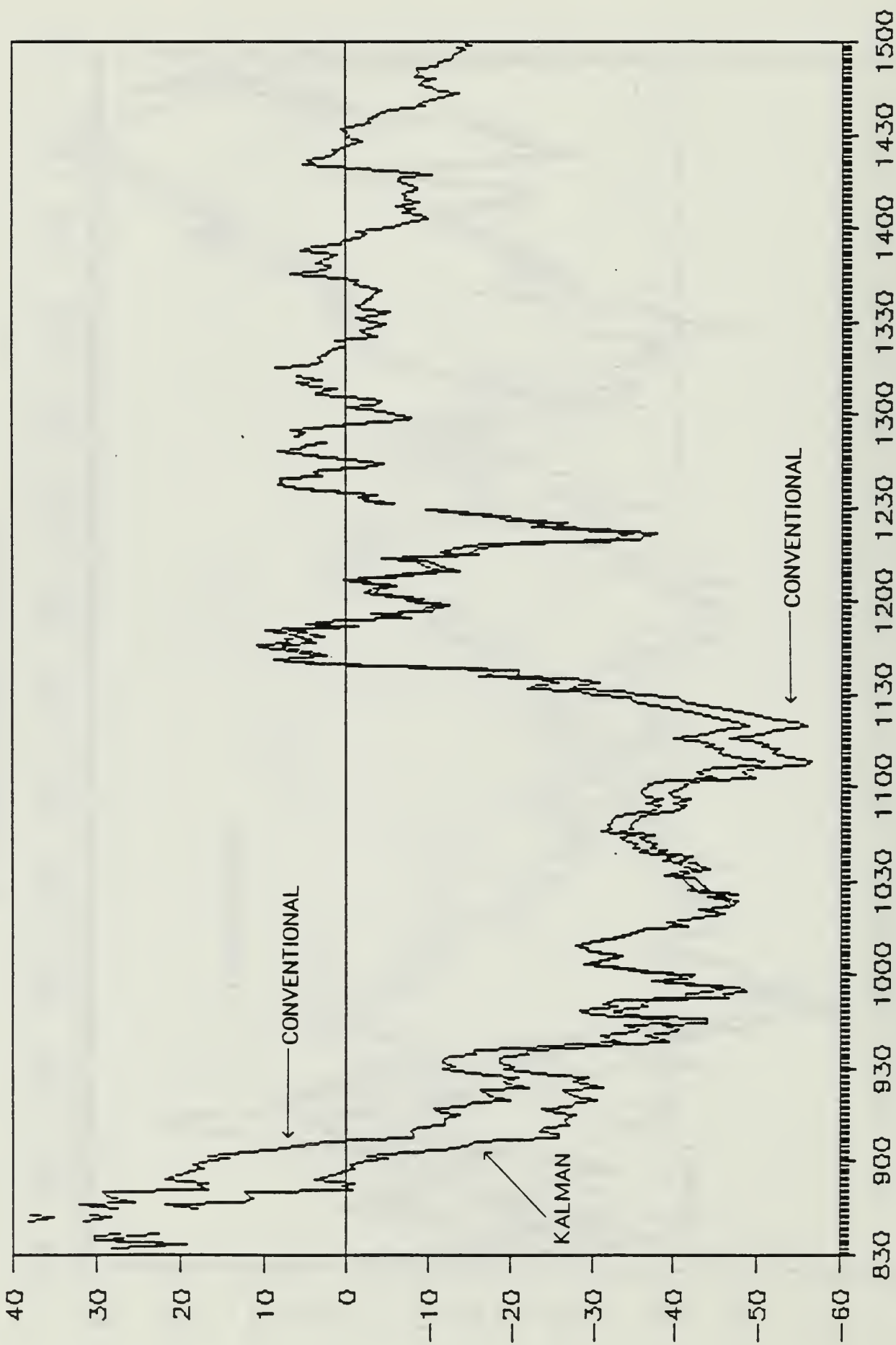
Figure 25   2—Sigma Confidence Intervals

October 19th, 1987
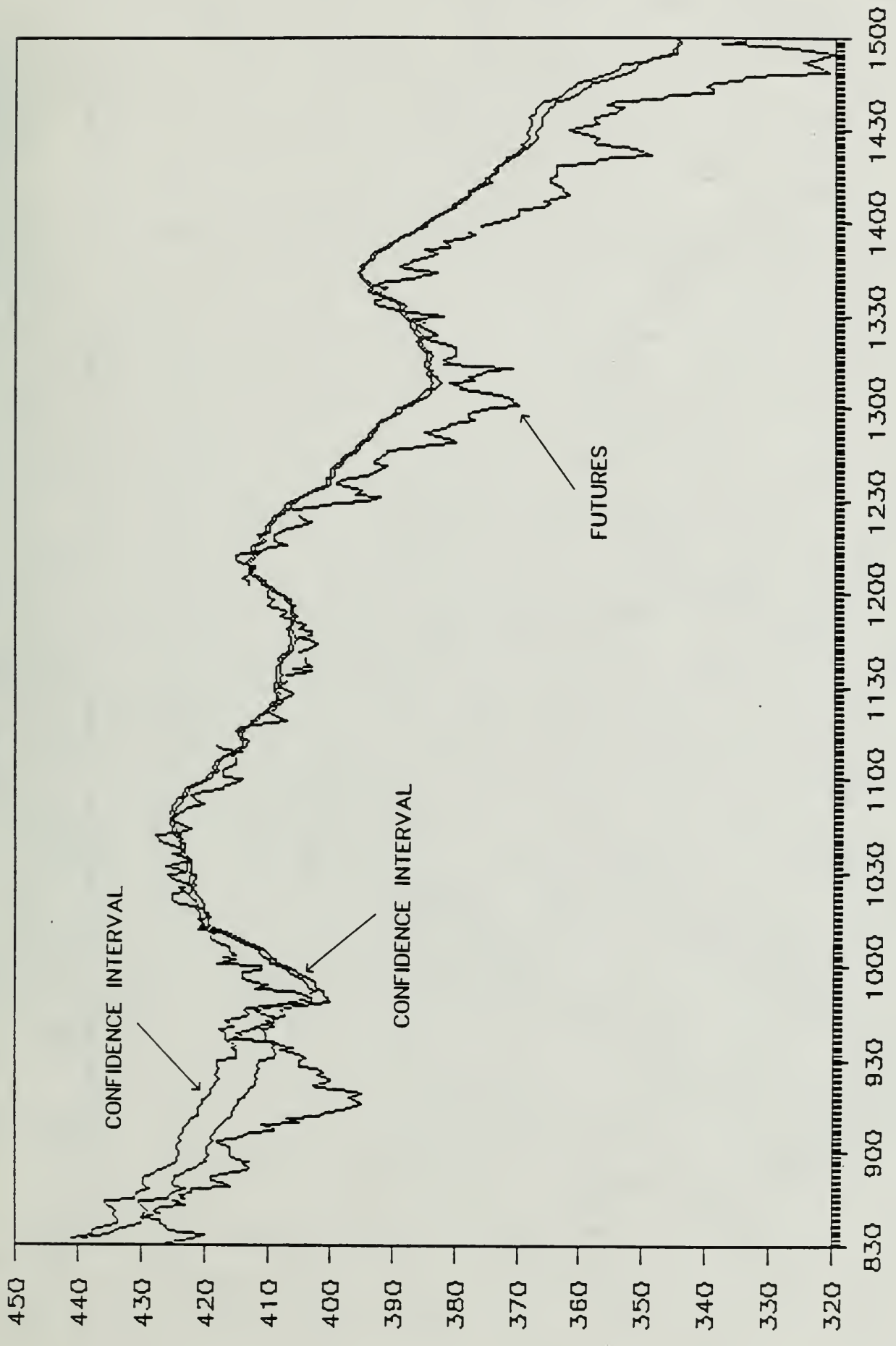
Figure 26 2-Sigma Confidence Intervals

October 20th, 1987